Estimation of extremes for Weibull-tail distributions in the presence of random censoring

Julien Worms, Rym Worms

Estimation of extremes for Weibull-tail distributions in the presence of random censoring

Julien Worms (1) & Rym Worms[1] (2)

(1) Université Paris-Saclay/Université de Versailles-Saint-Quentin-En-Yvelines
Laboratoire de Mathématiques de Versailles (CNRS UMR 8100),
F-78035 Versailles Cedex, France,
e-mail : julien.worms@uvsq.fr

(2) Université Paris-Est
Laboratoire d'Analyse et de Mathématiques Appliquées
(CNRS UMR 8050),
UPEMLV, UPEC, F-94010, Créteil, France,
e-mail : rym.worms@u-pec.fr

Abstract

The Weibull-tail class of distributions is a sub-class of the Gumbel extreme domain of attraction, and it has caught the attention of a number of researchers in the last decade, particularly concerning the estimation of the so-called Weibull-tail coefficient. In this paper, we propose an estimator of this Weibull-tail coefficient when the Weibull-tail distribution of interest is censored from the right by another Weibull-tail distribution: to the best of our knowledge, this is the first one proposed in this context. A corresponding estimator of extreme quantiles is also proposed. In both mild censoring and heavy censoring (in the tail) settings, asymptotic normality of these estimators is proved, and their finite sample behavior is presented via some simulations.

AMS Classification. Primary 62G32 ; Secondary 62N02

Keywords and phrases. Weibull-tail. Tail inference. Random censoring. Asymptotic representation.

[1]Correspondent author

1. Introduction

In recent years, the problem of studying extreme events and estimating extreme quantiles for randomly censored data has caught the attention of a growing number of researchers, due to the numerous applications which call for concrete solutions. Examples of such domains of application are non-life insurance, survival analysis, system or "material" reliability... Beirlant et al. (2007) and Einmahl et al. (2008) presented a general method for adapting estimators of the extreme value index in this censorship framework. Worms and Worms (2014), Beirlant et al. (2019) and Worms and Worms (2015) proposed a more survival analysis-oriented approach, the two first being restricted to the heavy tail case. Ndao et al. (2014), Ndao et al. (2016) and Stupfler (2016) extended the framework to data with covariate information. Beirlant et al. (2016) and Beirlant et al. (2018) proposed bias-reduced versions of two existing estimators. See also Brahimi et al. (2015), Brahimi et al. (2016) and Brahimi et al. (2018) for other papers on the subject.

However, a number of these works assume that the observed data come from heavy-tailed distributions (for both the sample of interest and the censoring sample), while many applications for which extreme events need to be studied do not exhibit a heavy-tail behavior, particularly in the survival analysis domain, where the censored data are lifetimes of patients or of animals, or time-to-failure of systems or items. For example, in Gomes and Neves (2011), the authors show that some larynx cancer or leucomia datasets do not exhibit a heavy right-tail.

We consider in this paper the Weibull-tail framework, where both the censored and censoring distributions have exponentially decreasing survival functions, driven by a coefficient defined a few lines below and called the Weibull-tail coefficient. This sub-class of the Gumbel max-domain of attraction has been the topic of a fair amount of papers in the extreme value analysis literature (Beirlant et al. (1995), Girard (2004,a), Gardes and Girard (2005), Diebolt et al. (2008), Goegebeur et al. (2010), to name just a few). But, to the best of our knowledge, all of them took place in the complete data setup. The present paper seems to be the first to propose an estimator of the Weibull-tail coefficient adapted to random censoring. As a corollary, a new estimator of extreme quantiles for light-tailed data will be studied.

Let us now detail the exact framework of this paper. We consider the observation of a sample of n independent couples $(Z_i, \delta_i)_{1 \leqslant i \leqslant n}$ where

$$Z_i = \min(X_i, C_i) \qquad \text{and} \quad \delta_i = \mathbb{I}_{X_i \leqslant C_i}. \tag{1}$$

In this definition, the i.i.d. samples $(X_i)_{i \leqslant n}$ and $(C_i)_{i \leqslant n}$, of respective continuous distribution functions F and G, are samples from the variable of interest X and of the censoring variable C, measured on n individual items (insurance claims, hospitalized patients, ...), but for each item or individual, only one of the two measurements (the lowest one) is observed. The variables X and C are supposed to be independent and we will suppose in this work that they are non-negative. We will denote by $Z_{1,n} \leqslant \ldots \leqslant Z_{i,n} \leqslant \ldots \leqslant Z_{n,n}$ the order statistics associated to the observed sample, and by $(\delta_{1,n}, \ldots, \delta_{n,n})$ the corresponding indicators of non-censorship.

The goal is to investigate the right-tail of F, and the main assumption of this paper is that, in the relations

$$\bar{F}(x) = 1 - F(x) = \exp(-\Lambda_F(x)) \quad \text{and} \quad \bar{G}(y) = 1 - G(y) = \exp(-\Lambda_G(y)), \tag{2}$$

the cumulative hazard functions Λ_F and Λ_G are semi-parametrically modeled by the relations

$$\Lambda_F(x) = x^{1/\theta_X} l_F(x) \quad \text{and} \quad \Lambda_G(y) = y^{1/\theta_C} l_G(y), \tag{3}$$

for some positive parameters θ_X and θ_C and slowly varying functions (at $+\infty$) l_F and l_G. This setup is the one where F and G are said to be Weibull-tailed, and θ_X and θ_C are the so-called Weibull-tail coefficients of F and G.

Our aim is to estimate the coefficient θ_X using the observed sample $(Z_i)_{i \leqslant n}$ and the observed non-censoring indicators $(\delta_i)_{i \leqslant n}$. Noting H the cumulative distribution function of the observable Z, and $\bar{H}(x) = 1 - H(x) = \mathbb{P}(Z > x)$, by independence of the samples X and C we have $\bar{H}(x) = \bar{F}(x)\bar{G}(x) = \exp(-\Lambda_H(x))$, where

$$\Lambda_H(x) = \Lambda_F(x) + \Lambda_G(x) = x^{1/\theta_X} l_F(x) + x^{1/\theta_C} l_G(x) = x^{1/\theta_Z} l_H(x), \quad \text{where} \quad \theta_Z = \min\{\theta_X, \theta_C\}$$

and l_H is a slowly varying function at infinity. More details on this function (and on other slowly varying functions) will be provided later in this paper.

The case where $\theta_X \leqslant \theta_C$ can be viewed as the case where the censoring tail is similar to, or heavier than, the tail of the variable X of interest, *i.e.* the censoring is expected to be moderate *in the tail* (more

details about this in a few lines). In this case, the Weibull-tail coefficient θ_Z of the data Z is equal to the Weibull-tail coefficient θ_X we wish to estimate, and so this would suggest that trying to define an estimator of θ_X adapted to censoring is a waste of time : however, as simulations show (see Section 5), not taking into account the censoring mechanism, by estimating θ_X by any non-adapted-to-censoring estimator of θ_Z based on the observed data Z_i, can often lead to an unreliable estimate of θ_X. So if theory suggests that the topic of adapting Weibull-tail estimation to random censoring sounds like a non-problem (in the mild censoring case), it turns out to be in practice an important issue which needs to be addressed.

The case where $\theta_C < \theta_X$ is the case where the tail of the censored variable X is heavier than the tail of the censored variable C, *i.e.* the censoring is expected to be strong in the tail. In this case, the Weibull-tail coefficient θ_Z of the observed data is no longer equal to the one of the original sample X, it is equal to θ_C : in this situation, an appropriate strategy needs to be developed, which is detailed below.

Moreover, in practice, it is difficult to know a priori the position of the Weibull-tail coefficient of X with respect to the one of C : the definition of our estimator of θ_X (see below) does not presume anything about this position (however, the rate of convergence and asymptotic variance will differ whether θ_X is lower than θ_C or not).

It is important to note that the position of θ_X with respect to θ_C has an important impact on the amount of censoring in the tail. As a matter of fact, Lemma 3 (in the Appendix) states that the *ultimate* probability of non-censoring in the tail (limit of $\mathbb{P}(\delta = 1|Z = z)$ for $z \to \infty$, denoted by p later on) turns out to be equal to 1 when $\theta_X < \theta_C$, to 0 when $\theta_X > \theta_C$, and to a constant between 0 and 1 when $\theta_X = \theta_C$. It is however important to remember that this is an asymptotic value, and in practice, for finite sample sizes, things are less clear-cut (the simulation Section 5 illustrates this). Moreover, other characteristics of the underlying distributions (for instance, position or scale parameters) may have a non-negligeable impact on the proportion of censoring, even in the tail : this delicate topic should deserve more attention in subsequent works. In the sequel, the situation $\theta_X \leqslant \theta_C$ will nonetheless be referred to as the "mild censoring" setting, opposed to the "strong censoring" setting when $\theta_C < \theta_X$.

Let us now explain how our estimator is defined. In the non-censored case (*i.e.* if we could observe the original data values $X_1, \ldots, X_n$), the usual starting point for designing estimators of the Weibull-tail coefficient is to note that, by slow variation of the function l_F defined in (3), we have, for t large and any $x > 1$, the approximation $\theta_X \log(\Lambda_F(tx)/\Lambda(t)) \simeq \log(x)$. Therefore, for some value $k = k_n$ (the number of top order statistics used in the estimation) to be chosen, considering $t = X_{n-k,n}$ and $x = X_{n-j+1,n}/X_{n-k,n}$ for every $1 \leqslant j \leqslant k$ in the above formula leads, after summation, to

$$\theta_X \simeq \frac{\sum_{j=1}^{k}(\log(X_{n-j+1,n}) - \log(X_{n-k,n}))}{\sum_{j=1}^{k}(\log(\Lambda_F(X_{n-j+1,n})) - \log(\Lambda_F(X_{n-k,n})))}, \tag{4}$$

where $X_{1,n}, \ldots, X_{n,n}$ are the order statistics. As was initiated in Beirlant et al. (1995) and developed in Girard (2004,a), this suggests define an estimator of θ_X in the complete data case by

$$\hat{\theta}_X^{(complete)} = \frac{\sum_{j=1}^{k}(\log(X_{n-j+1,n}) - \log(X_{n-k,n}))}{\sum_{j=1}^{k}(\log(\log(n/j)) - \log(\log(n/k)))}, \tag{5}$$

because $\log \Lambda_F(x) = \log(-\log \bar{F}(x))$ and $\bar{F}$ evaluated at some order statistic $X_{n-j+1,n}$ can be naturally estimated by j/n. However, in the censored setup, the observed variables are the Z_i, which are associated with Λ_H, and not with Λ_F : therefore the previous trick that led to the deterministic denominator $\sum_{j=1}^{k}(\log(\log(n/j)) - \log(\log(n/k)))$ cannot be used. Our proposition in the censored context is simply to replace, in formula (4), the X's with the observed Z's, and to estimate the function Λ_F by its Nelson-Aalen estimator

$$\hat{\Lambda}_{nF}(x) = \sum_{Z_{i,n} \leqslant x} \frac{\delta_{i,n}}{n - i + 1}. \tag{6}$$

This leads to our proposition for estimating θ_X in the censored setup :

$$\hat{\theta}_{X,k} = \frac{\sum_{j=1}^{k}(\log Z_{n-j+1,n} - \log Z_{n-k,n})}{\sum_{j=1}^{k}\left(\log \hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \log \hat{\Lambda}_{nF}(Z_{n-k,n})\right)}. \tag{7}$$

In contrast with the estimator of θ_X in the complete data framework (and with a number of its variants), our estimator has a random denominator, which behavior will turn out to be closely related to that of the numerator.

Note that our estimator can be written as the ratio

$$\hat{\theta}_{X,k} = \frac{\hat{\theta}_{Z,k}}{RL_n},$$

where

$$\hat{\theta}_{Z,k} = \frac{\sum_{j=1}^{k} (\log Z_{n-j+1,n} - \log Z_{n-k,n})}{\sum_{j=1}^{k} (\log(\log(n/j)) - \log(\log(n/k)))} \quad \text{and} \quad RL_n = \frac{\sum_{j=1}^{k} \left(\log \hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \log \hat{\Lambda}_{nF}(Z_{n-k,n}) \right)}{\sum_{j=1}^{k} (\log(\log(n/j)) - \log(\log(n/k)))}. \tag{8}$$

The numerator $\hat{\theta}_{Z,k}$ estimates the Weibull-tail coefficient θ_Z of the observed Z_i (see Theorem 1 in Girard (2004,a)). As far as only consistency is studied, it is possible to prove consistency of $\hat{\theta}_{X,k}$ by proving that the denominator RL_n converges to some crucial value $a = \theta_Z/\theta_X$, which is equal to 1 in the mild censoring cases ($\theta_X < \theta_X$ or $\theta_X = \theta_C$), and is lower than 1 in the strong censoring cases ($\theta_C < \theta_X$). This is in fact deduced from the proof of Theorem 1 and is stated later in this paper (Corollary 1). However, to establish the asymptotic normality of our estimator, things are more complicated and we invite the interested reader to have a look at the start of the proof of Theorem 1 in Section 3.

Anyway, it is interesting to note that, in a Weibull-tail situation, a possible correction for censoring could be to divide an existing estimator of θ_Z in the complete data setup, by this statistic RL_n that somehow incorporates the censoring information of the data. This is similar to what is proposed in Beirlant et al. (2007) and Einmahl et al. (2008) for adapting estimators of the extreme value index to the censoring situation, namely the now well-known "division by the proportion of non-censoring in the tail" strategy. However, note that we do not know whether this strategy still leads to valuable estimators when applied to other estimators of θ_Z than the basic estimator $\hat{\theta}_{Z,k}$ defined in (8).

Concerning now the estimation of extreme quantiles for Weibull-tail censored data, we propose to consider, for any given small probability $p_n < 1/n$, the natural estimator of $x_{p_n} = \bar{F}^{-}(p_n)$ defined by

$$\hat{x}_{p_n} := Z_{n-k,n} \left(\frac{-\log p_n}{\Lambda_{n,F}(Z_{n-k,n})} \right)^{\hat{\theta}_{X,k}}. \tag{9}$$

This definition comes from the application, to the values $x = x_{p_n}/Z_{n-k,n}$ and $t = Z_{n-k,n}$, of the approximation $x \simeq (\Lambda_F(tx)/\Lambda_F(t))^{\theta_X} = (-\log \bar{F}(tx)/\Lambda_F(t))^{\theta_X}$, valid for t large and any $x > 1$.

Before going into the details of our results, we indicate here that in this work it is assumed that the slowly varying functions l_F and l_G, defined in (3), both satisfy the classical second order condition SR2 :

$$\forall x > 0, \frac{\frac{l_F(tx)}{l_F(t)} - 1}{b_F(t)} \to K_{\rho_F}(x), \quad \text{and} \quad \frac{\frac{l_G(tx)}{l_G(t)} - 1}{b_G(t)} \to K_{\rho_G}(x), \text{ as } t \to +\infty, \text{ where } K_\rho(x) := \frac{x^\rho - 1}{\rho}, \tag{10}$$

for some negative constants ρ_F and ρ_G, with rate functions b_F and b_G having constant sign at $+\infty$ and satisfying $|b_F| \in RV_{\rho_F}$ and $|b_G| \in RV_{\rho_G}$ (RV_ρ stands for regular variation, at $+\infty$, with index ρ).

Our paper is organized as follows: in Section 2, we state the asymptotic normality result for $\hat{\theta}_n$ and $\hat{x}_{p_n}$. Section 3 and 4 is devoted to the proofs. Important lemmas and technical aspects of the proofs are postponed to the Appendix. In Section 5, we discuss the finite sample behavior of our new estimators.

2. Results

Let us first introduce the following important quantities

$$a = \frac{\theta_Z}{\theta_X} = \begin{cases} 1 & \text{if } \theta_X \leqslant \theta_C, \\ \theta_C/\theta_X \in]0,1[& \text{if } \theta_X > \theta_C, \end{cases}, \; b = \frac{1-a}{2} \text{ and } d = \frac{\theta_X}{\theta_C}.$$

Such definitions will be useful to state results in a general way, without having to discuss whether we are in a mild or in a strong (ultimate) censoring setting.

We have seen in the introduction that the cumulative hazard function Λ_H of Z is regularly varying of order $1/\theta_Z$. Setting Λ_H^- for the generalized inverse of Λ_H, we then have

$$\Lambda_H^-(x) = x^{\theta_Z} l(x) \quad \text{and} \quad \Lambda_F \circ \Lambda_H^-(x) = x^a \tilde{l}(x),$$

where l and $\tilde{l}$ are slowly varying at infinity. The second formula is important in our setting since, by definition of our estimator $\hat{\theta}_{X,k}$, we will have to deal with the quantities $\Lambda_F(Z_{n-j+1,n})$.

By Lemma 2 stated in the Appendix, and in its subsequent remark, we know that under assumption (10), there exist positive constants c_F, c_G, c and $\tilde{c}$ such that, for $x > 0$,

$$l_F(x) = c_F(1 - x^{\rho_F} v_F(x)) \quad \text{and} \quad l_G(x) = c_G(1 - x^{\rho_G} v_G(x)),$$
$$l(x) = c(1 - x^{\rho} v(x)) \quad \text{and} \quad \tilde{l}(x) = \tilde{c}(1 - x^{\tilde{\rho}} \tilde{v}(x)),$$

where $|v_F|$, $|v_G|$, $|v|$ and $|\tilde{v}|$ are slowly varying functions at infinity. Therefore, the functions l and $\tilde{l}$ satisfy an SR2 condition, with negative second order parameter respectively denoted by ρ and $\tilde{\rho}$ (exact expressions are provided in Lemma 2) and respective rate function B and $\tilde{B}$, with constant sign at $+\infty$ and their absolute value being regularly varying with respective index ρ and $\tilde{\rho}$. However, for technical reasons, we need to assume the following stronger conditions noted $R_l(B, \rho)$ and $R_{\tilde{l}}(\tilde{B}, \tilde{\rho})$, defined by :

$\boldsymbol{R_\ell(B, \rho)}$: There exists a constant $\rho < 0$ and a rate function B satisfying $B(x) \to 0$, as $x \to +\infty$, such that for all $\epsilon > 0$ and $A > 1$, we have

$$\sup_{\lambda \in [1, A]} \left| \frac{\ell(\lambda x)/\ell(x) - 1}{B(x) K_\rho(\lambda)} - 1 \right| \leqslant \epsilon, \text{ for } x \text{ sufficiently large },$$

with $|B|$ being necessarily regularly varying, at $+\infty$, with index ρ.

In order to obtain the asymptotic normality of our estimator, we need the sequence (k_n) (number of top order statistics to use) to satisfy the following conditions (we will note $k = k_n$ from now on):

$H_1 : k \to +\infty$, $k/n \to 0$, $\frac{\log k}{\log n} \to 0$, as $n \to +\infty$,

and, depending on the censoring strength in the tail, introducing the important notation $L_{nk} = \log(n/k)$,

$$H_2 : \theta_X < \theta_C \text{ and } \begin{cases} (i)\ \sqrt{k}\, B(L_{nk}) \to \alpha \\ (ii)\ \sqrt{k}\, \tilde{B}(L_{nk}) \to \tilde{\alpha} \\ (iii)\ \sqrt{k}\, L_{nk}^{d-1} \to \alpha' \end{cases}$$

$$H_3 : \theta_X = \theta_C \text{ and } \begin{cases} (i)\ \exists \delta > 0, \sqrt{k}\, L_{nk}^{\rho+\delta} \to 0, \text{ where } \rho = \max(\theta_Z \rho_F, \theta_Z \rho_G) \\ (ii) \sqrt{k}\, L_{nk}^{-1} \to 0 \end{cases}$$

$$H_4 : \theta_X > \theta_C \text{ and } \begin{cases} (i)\ \sqrt{k} L_{nk}^{-b} \to +\infty \\ (ii)\ \exists \delta > 0, \sqrt{k}\, L_{nk}^{-b+\tilde{\rho}+\delta} \to 0, \text{ where } \tilde{\rho} = \max(\theta_Z \rho_F, \theta_Z \rho_G, a - 1) \\ (iii)\ \exists \delta > 0, \sqrt{k}\, L_{nk}^{-(1+b)/2+\delta} \to 0 \\ (iv)\ \sqrt{k}\, L_{nk}^{-b-a} \to 0 \end{cases}$$

See Remark 2 below for a discussion on those conditions. Our main result is the following theorem.

Theorem 1. *Let conditions* (2), (3) *and* (10) *hold, as well as* $R_l(B, \rho)$ *and* $R_{\tilde{l}}(\tilde{B}, \tilde{\rho})$. *We assume further that* (k_n) *satisfies conditions* H_1 *and either* H_2, H_3 *or* H_4. *We then have, as* $n \to \infty$,

$$\sqrt{k} L_{nk}^{-b} (\hat{\theta}_{X,k} - \theta_X) \xrightarrow{d} N\left(m, \frac{\theta_X^2}{a\tilde{c}}\right),$$

where

$$\tilde{c} = \begin{cases} 1 & \text{if } \theta_X < \theta_C \\ c_F/(c_F + c_G) & \text{if } \theta_X = \theta_C \\ c_G^{-a} c_F & \text{if } \theta_X > \theta_C \end{cases} \quad \text{and} \quad m = \begin{cases} \alpha + \dfrac{\tilde{\alpha}\theta_X}{\rho} + \dfrac{\theta_X^2}{\theta_C} \dfrac{c_G}{c_F^d} \alpha' & \text{if } \theta_X < \theta_C, \\ 0 & \text{if } \theta_X \geqslant \theta_C. \end{cases}$$

Remark 1. *When* $\theta_X \leqslant \theta_C$, b *is equal to* 0 *and thus, the rate of convergence* $\sqrt{k} L_{nk}^{-b}$ *is the same as in the non-censored case. It is slower when* $\theta_X > \theta_C$. *The asymptotic variance* $\theta_X^2/(a\tilde{c})$ *equals* θ_X^2 *when* $\theta_X < \theta_C$ *(i.e. the same asymptotic variance as in the non-censored situation), and is larger than* θ_X^2 *when* $\theta_X = \theta_C$ *; nothing can be said in general about its position with respect to* θ_X^2 *when* $\theta_X > \theta_C$.

Remark 2. *When* $\theta_X < \theta_C$, *rate functions* $|B|$ *and* $|\tilde{B}|$ *appearing in assumptions* $R_l(B, \rho)$ *and* $R_{\tilde{l}}(\tilde{B}, \tilde{\rho})$ *are regularly varying of same order* $\rho = \tilde{\rho} = \max(\theta_X \rho_F, d - 1)$ *(see Lemma 2 in the Appendix), therefore, either* $\rho = d - 1$ *and thus conditions* H_2 $(i), (ii), (iii)$ *essentially involve the same rate condition on* k_n, *or* $\rho > d - 1$ *and thus condition* $H_2(i)$ *or* (ii) *implies condition* $H_2(iii)$, *with* $\alpha' = 0$.
When $\theta_X = \theta_C$, *if* $\rho \geqslant -1$, *then condition* $H_3(i)$ *implies condition* $H_3(ii)$, *and if* $\rho < -1$, *the implication is reversed. When* $\theta_X \geqslant \theta_C$, *only one of the conditions* H_4 $(ii), (iii), (iv)$ *remains, depending on the position of*

a and $\tilde{\rho}$.
Moreover, conditions $H_2(i)$ and $H_2(ii)$, involving the regularly varying functions B and $\tilde{B}$, do not appear in the cases $\theta_X \geqslant \theta_C$, because they are consequences of $H_3(i)$ or $H_4(ii)$, with $\alpha = \alpha' = 0$, necessarily.

Before stating the asymptotic normality of our extreme quantile estimator $\hat{x}_{p_n}$ defined in (9), we need to introduce the following additional conditions (as $n \to \infty$) :
$H_1' : \frac{\log L_{nk}}{\log\log(1/p_n)} \to 0,$
$H_2(iv) : \sqrt{k}\, L_{nk}^{\theta_X \rho_F} \to \alpha''.$

Theorem 2. *Let conditions (2), (3) and (10) hold, as well as $R_l(B,\rho)$ and $R_{\tilde{l}}(\tilde{B},\tilde{\rho})$. We assume further that (k_n) satisfies conditions H_1, H_1' and either H_2, H_3 or H_4. We then have, as $n \to \infty$,*

$$\frac{\sqrt{k} L_{nk}^{-b}}{\log\log(1/p_n)} (\log \hat{x}_{p_n} - \log x_{p_n}) \xrightarrow{d} N\left(m, \frac{\theta_X^2}{a\tilde{c}}\right).$$

3. Proof of Theorem 1

Remind that

$$\hat{\theta}_{X,k} = \frac{\frac{1}{k}\sum_{j=1}^{k} (\log Z_{n-j+1,n} - \log Z_{n-k,n})}{\frac{1}{k}\sum_{j=1}^{k} \left(\log \hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \log \hat{\Lambda}_{nF}(Z_{n-k,n})\right)}.$$

Introducing $E_1, \ldots, E_n$ n independent standard exponential random variables, such that $Z_i = \Lambda_H^-(E_i)$, we have, since $\Lambda_H^-(x) = x^{\theta_Z} l(x)$ and $\Lambda_F \circ \Lambda_H^-(x) = x^a \tilde{l}(x)$ with l and $\tilde{l}$ slowly varying at infinity,

$$\log Z_{n-j+1,n} - \log Z_{n-k,n} = \theta_Z \log\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right) + \log\left(\frac{l(E_{n-j+1,n})}{l(E_{n-k,n})}\right) \tag{11}$$

$$\log \Lambda_F(Z_{n-j+1,n}) - \log \Lambda_F(Z_{n-k,n}) = a \log\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right) + \log\left(\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})}\right), \tag{12}$$

Now, let

$$M_n = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right),$$

and

$$\Delta_n = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{\hat{\Lambda}_{nF}(Z_{n-j+1,n})}{\Lambda_F(Z_{n-j+1,n})} \frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}\right). \tag{13}$$

Since the denominator in the expression for $\hat{\theta}_{X,k}$ above equals

$$\frac{1}{k}\sum_{j=1}^{k} \left(\log \hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \log \hat{\Lambda}_{nF}(Z_{n-k,n})\right) = \frac{1}{k}\sum_{j=1}^{k} \log \Lambda_F(Z_{n-j+1,n}) - \log \Lambda_F(Z_{n-k,n}) + \Delta_n,$$

we obtain, using (11), (12) and relation $\theta_X = \theta_Z/a$,

$$\begin{aligned}
\hat{\theta}_{X,k} - \theta_X &= \frac{\theta_Z M_n + R_{n,l}}{a M_n + R_{n,\tilde{l}} + \Delta_n} - \theta_X \\
&= \theta_X \frac{\theta_X^{-1} R_{n,l} - R_{n,\tilde{l}} - \Delta_n}{a M_n + R_{n,\tilde{l}} + \Delta_n} \\
&= -\frac{\theta_X}{a}\Delta_n \left(M_n + a^{-1} R_{n,\tilde{l}} + a^{-1}\Delta_n\right)^{-1} + \frac{R_{n,l} - \theta_X R_{n,\tilde{l}}}{a M_n + R_{n,\tilde{l}} + \Delta_n},
\end{aligned}$$

where

$$R_{n,l} = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{l(E_{n-j+1,n})}{l(E_{n-k,n})}\right) \quad \text{and} \quad R_{n,\tilde{l}} = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})}\right). \tag{14}$$

We thus have the following representation, which shows that the behavior of the estimation error is essentially based on the behavior of the statistic Δ_n :

$$\sqrt{k}L_{nk}^{-b}\left(\hat{\theta}_{X,k}-\theta_X\right)=\left(-\frac{\theta_X}{a}\right)\sqrt{k}L_{nk}^{1-b}\Delta_n D_n^{-1}+\left(\sqrt{k}L_{nk}^{1-b}R_{n,l}-\theta_X\sqrt{k}L_{nk}^{1-b}R_{n,\tilde{l}}\right)(aD_n)^{-1}$$

where the denominator $D_n = L_{nk}M_n + a^{-1}L_{nk}R_{n,\tilde{l}} + a^{-1}L_{nk}\Delta_n$ will turn out to converge to 1. It is now clear that the proof of Theorem 1 then follows from the combination of the following three propositions, the first one being the most important and the longest to establish. These propositions are proved in the next three subsections.

Proposition 1. *Under the conditions of Theorem 1 we have, as n tends to infinity,*

$$\Delta_n \stackrel{d}{=} \frac{1+o_{\mathbb{P}}(1)}{L_{nk}}\left(\left(L_{nk}^{1-a}\frac{\hat{p}_k}{\tilde{c}}-a\right)-a\left(\bar{E}_n-1\right)\right)+o_{\mathbb{P}}(k^{-1/2}L_{nk}^{b-1}) \tag{15}$$

and

$$\sqrt{k}L_{nk}^{1-b}\Delta_n \xrightarrow{d} N\left(m_\Delta, \frac{a}{\tilde{c}}\right),$$

where $\bar{E}_n = \frac{1}{k}\sum_{i=1}^{k} E_i$ *(sample mean of standard exponential variables), and*

$$\hat{p}_k := \frac{1}{k}\sum_{j=1}^{k}\delta_{n-j+1,n} \qquad \textit{and} \qquad m_\Delta = \begin{cases} -\tilde{\alpha}\left(1+\dfrac{1}{\rho}\right)-\dfrac{\theta_X}{\theta_C}\dfrac{c_G}{c_F^d}\alpha' & \textit{if } \theta_X<\theta_C, \\ 0 & \textit{if } \theta_X \geqslant \theta_C. \end{cases}$$

Please note that the exponential variables E_i appearing in the statement of Proposition 1 are not the same as those introduced at the beginning of this Section.

Proposition 2. *Under the conditions of Theorem 1 we have, as n tends to infinity,*

$$\sqrt{k}L_{nk}^{1-b}R_{n,l}\xrightarrow{\mathbb{P}}\begin{cases}\alpha & \textit{if } \theta_X<\theta_C,\\ 0 & \textit{if } \theta_X\geqslant\theta_C,\end{cases} \quad \textit{and} \quad \sqrt{k}L_{nk}^{1-b}R_{n,\tilde{l}}\xrightarrow{\mathbb{P}}\begin{cases}\tilde{\alpha} & \textit{if } \theta_X<\theta_C,\\ 0 & \textit{if } \theta_X\geqslant\theta_C.\end{cases}$$

Proposition 3. *Under condition H_1, we have $L_{nk}M_n \xrightarrow{\mathbb{P}} 1$, as n tends to infinity.*

Remark 3. *First, remind that $a = 1$ and $\tilde{c} = 1$ when $\theta_X < \theta_C$. Let us highlight that the convergence in distribution of $\sqrt{k}L_{nk}^{1-b}\Delta_n$ stated in Proposition 1 comes from the confrontation between the two terms appearing in the representation* (15) *of Δ_n : the term in $\hat{p}_k$ and the term involving the exponential sample mean. The convergence in distribution of the term involving $\hat{p}_k$ is detailed in Lemma 1 in Subsection 3.1; this will be the leading term only when $\theta_X > \theta_C$ (in this setting, the constant b is positive and thus the exponential term vanishes). When $\theta_X < \theta_C$, it will only generate a possible bias, and when $\theta_X = \theta_C$ it participates to the asymptotic normality along with the exponential term.*

The following corollary is then stated, concerning the statistic RL_n defined in equation (8) and discussed thereafter. Note that this corollary certainly holds with weaker conditions.

Corollary 1. *Under the conditions of Theorem 1, as $n \to \infty$, we have $RL_n \xrightarrow{\mathbb{P}} a$.*

Its proof is short, so we will provide it here. With the same notations as in the previous page, we have readily

$$RL_n = \left(\frac{1}{k}\sum_{j=1}^{k}\log\log(n/j)-\log\log(n/k)\right)^{-1}\frac{1}{L_{nk}}(aL_{nk}M_n+L_{nk}R_{n,l}+L_{nk}\Delta_n),$$

where the mean inside the large brackets is equivalent to $1/L_{nk}$ (see Girard (2004,b) formula (15), for a proof). The proof of Corollary 1 thus follows from Propositions 1, 2 and 3.

3.1. Proof of Proposition 1

Starting from the definition of Δ_n in (13), we introduce the first remainder term $R_{1,k}^{(\Delta)}$ by writing

$$\begin{aligned}\Delta_n &= \frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{\hat{\Lambda}_{nF}(Z_{n-j+1,n})}{\Lambda_F(Z_{n-j+1,n})}\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}\right)\\ &= \frac{1}{k}\sum_{j=1}^{k}\left(\frac{\hat{\Lambda}_{nF}(Z_{n-j+1,n})}{\Lambda_F(Z_{n-j+1,n})}\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}-1\right)+R_{1,k}^{(\Delta)}.\end{aligned}$$

Now, using the definition of $\hat{\Lambda}_{nF}$ in (6), we obtain

$$\frac{1}{k}\sum_{j=1}^{k}\left(\hat{\Lambda}_{nF}(Z_{n-j+1,n})-\hat{\Lambda}_{nF}(Z_{n-k,n})\right)=\frac{1}{k}\sum_{j=1}^{k}\sum_{i=j}^{k}\frac{\delta_{n-j+1,n}}{j}=\frac{1}{k}\sum_{j=1}^{k}\delta_{n-j+1,n}=\hat{p}_k.$$

Hence, it can easily be checked that

$$\frac{\hat{\Lambda}_{nF}(Z_{n-k,n})}{\Lambda_F(Z_{n-k,n})}\left(\Delta_n-R_{1,k}^{(\Delta)}\right) \;=\; \frac{\hat{p}_k}{\Lambda_F(Z_{n-k,n})}-\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\Lambda_F(Z_{n-j+1,n})}{\Lambda_F(Z_{n-k,n})}-1\right)+R_{2,k}^{(\Delta)},$$

where

$$R_{2,k}^{(\Delta)}=\frac{1}{\Lambda_F(Z_{n-k,n})}\frac{1}{k}\sum_{j=1}^{k}\left(\hat{\Lambda}_{nF}(Z_{n-j+1,n})-\Lambda_F(Z_{n-j+1,n})\right)\left(\frac{\Lambda_F(Z_{n-k,n})}{\Lambda_F(Z_{n-j+1,n})}-1\right).$$

Since, $\forall\ 1\leqslant j\leqslant k+1$, $\Lambda_F(Z_{n-j+1,n})=(\Lambda_F\circ\Lambda_H^{-})(E_{n-j+1,n})=E_{n-j+1,n}^{a}\tilde{l}(E_{n-j+1,n})$, where $\tilde{l}$ is slowly varying and tends to $\tilde{c}$ at infinity (cf Lemma 2, in the Appendix), then

$$\frac{\Lambda_F(Z_{n-j+1,n})}{\Lambda_F(Z_{n-k,n})}-1=\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right)^{a}\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})}-1=\left(\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right)^{a}-1\right)+\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right)^{a}\left(\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})}-1\right),$$

and, introducing $(\tilde{E}_1,\ldots,\tilde{E}_k)$ k independent standard exponential random variable such that, according to Lemma 4, $(E_{n-j+1,n}-E_{n-k,n})_{1\leqslant j\leqslant k}\stackrel{d}{=}(\tilde{E}_{k,k},\ldots,\tilde{E}_{1,k})$, we can write

$$\frac{\hat{\Lambda}_{nF}(Z_{n-k,n})}{\Lambda_F(Z_{n-k,n})}\left(\Delta_n-R_{1,k}^{(\Delta)}\right)\stackrel{d}{=}\frac{\hat{p}_k}{\tilde{c}E_{n-k,n}^{a}}+R_{3,k}^{(\Delta)}-\frac{1}{k}\sum_{j=1}^{k}\left(a\frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}}\right)+R_{4,k}^{(\Delta)}+R_{5,k}^{(\Delta)}+R_{2,k}^{(\Delta)},$$

where

$$\begin{aligned}
R_{3,k}^{(\Delta)} &= \frac{\hat{p}_k}{E_{n-k,n}^{a}}\left(\frac{1}{\tilde{l}(E_{n-k,n})}-\frac{1}{\tilde{c}}\right)\\
R_{4,k}^{(\Delta)} &= -\frac{1}{k}\sum_{j=1}^{k}\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right)^{a}\left(\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})}-1\right)\\
R_{5,k}^{(\Delta)} &= -\frac{1}{k}\sum_{j=1}^{k}\left\{\left(\left(1+\frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}}\right)^{a}-1\right)-a\frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}}\right\}.
\end{aligned}$$

Let us summarize :

$$\Delta_n\stackrel{d}{=}\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}\left(\left(\frac{\hat{p}_k}{\tilde{c}E_{n-k,n}^{a}}-\frac{a}{E_{n-k,n}}\frac{1}{k}\sum_{j=1}^{k}\tilde{E}_j\right)+\sum_{i=2}^{5}R_{i,k}^{(\Delta)}\right)+R_{1,k}^{(\Delta)}.$$

But

$$\frac{\hat{p}_k}{\tilde{c}E_{n-k,n}^{a}}-\frac{a}{E_{n-k,n}}\frac{1}{k}\sum_{j=1}^{k}\tilde{E}_j=\frac{1}{E_{n-k,n}}\left(\left(L_{nk}^{1-a}\frac{\hat{p}_k}{\tilde{c}}-a\right)-a\left(\bar{E}_n-1\right)\right)+R_{6,k}^{(\Delta)},$$

where $\bar{E}_n=\frac{1}{k}\sum_{j=1}^{k}\tilde{E}_j$ and

$$R_{6,k}^{(\Delta)}=\frac{\hat{p}_k}{\tilde{c}E_{n-k,n}}\left(E_{n-k,n}^{1-a}-L_{nk}^{1-a}\right).$$

Finally,

$$\Delta_n\stackrel{d}{=}\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}\left(\frac{1}{E_{n-k,n}}\left(\left(L_{nk}^{1-a}\frac{\hat{p}_k}{\tilde{c}}-a\right)-a\left(\bar{E}_n-1\right)\right)+\sum_{i=2}^{6}R_{i,k}^{(\Delta)}\right)+R_{1,k}^{(\Delta)}. \tag{16}$$

We shall show, in Lemma 7 in the Appendix, that $\sqrt{k}L_{nk}^{1-b}\sum_{i=1}^{6}R_{i,k}^{(\Delta)}$ tends to constant. Moreover, we have $\sqrt{k}\left(\bar{E}_n-1\right)\xrightarrow{d}N(0,1)$, and, according to Lemmas 5 and 6, both $\frac{L_{nk}}{E_{n-k,n}}$ and $\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}$ tend to 1 as $n\to+\infty$. Hence

$$\sqrt{k}L_{nk}^{1-b}\Delta_n\stackrel{d}{=}(1+o_{\mathbb{P}}(1))\left(D_n-a\sqrt{k}L_{nk}^{-b}\left(\bar{E}_n-1\right)\right)+(1+o_{\mathbb{P}}(1))\sum_{i=1}^{6}R_{i,k}^{(\Delta)}, \tag{17}$$

where

$$D_n = \sqrt{k} L_{nk}^{-b} \left(L_{nk}^{1-a} \frac{\hat{p}_k}{\tilde{c}} - a \right), \text{ with } b = (1-a)/2.$$

It remains to study the behavior D_n, which is done in the following Lemma.

Lemma 1. *Under the assumptions of Theorem 1, we have, as $n \to +\infty$:*

1. *If $\theta_X < \theta_C$, then $D_n = \sqrt{k}(\hat{p}_k - 1) \xrightarrow{\mathbb{P}} -\frac{\theta_X}{\theta_C} \frac{c_G}{c_F^d} \alpha'$.*
2. *If $\theta_X = \theta_C$, then $D_n = \sqrt{k} \left(\frac{\hat{p}_k}{p} - 1 \right) \xrightarrow{d} N\left(0, \frac{1-p}{p}\right)$, where $p = \tilde{c} = \frac{c_F}{c_F + c_G}$.*
3. *If $\theta_X > \theta_C$ (hence $a < 1$ and $b \in]0, 1/2[$), then $D_n \xrightarrow{d} N(0, a/\tilde{c})$.*

Remark 4. *Lemma 1 shows, in particular, that the proportion of non-censored data in the tail $\hat{p}_k$ tends to 1, if $\theta_X < \theta_C$, to $p = \tilde{c} = \frac{c_F}{c_F + c_G}$, if $\theta_X = \theta_C$ and to 0 (with rate L_{nk}^{a-1}) if $\theta_X > \theta_C$. This has to be linked to the result of Lemma 3 (see the Appendix) concerning the limit of the function $p(\cdot)$ defined below.*

When $\theta_X < \theta_C$, Lemma 1 states that D_n converges to a constant : hence, via Lemma 7, the leading term in (17) is $\sqrt{k} L_{nk}^{-b} \left(\bar{E}_n - 1 \right) = \sqrt{k} \left(\bar{E}_n - 1 \right) \xrightarrow{d} N(0,1)$, and we thus obtain as desired $\sqrt{k} L_{nk}^{1-b} \Delta_n \xrightarrow{d} N(m_\Delta, 1)$ via Lemma 7, where m_Δ is defined in the statement of Proposition 1. When $\theta_X = \theta_C$, the constant b is still equal to 0 and both D_n and $\sqrt{k} \left(\bar{E}_n - 1 \right)$ (which are independent) take part into the asymptotic normality of Δ_n, with $D_n - a\sqrt{k} \left(\bar{E}_n - 1 \right) \xrightarrow{d} N(0, \sigma_\Delta^2)$ in relation (17), where $\sigma_\Delta^2 = \frac{1-p}{p} + a^2 = \frac{1}{\tilde{c}}$. Thus, we obtain $\sqrt{k} L_{nk}^{1-b} \Delta_n \xrightarrow{d} N(0, \frac{1}{\tilde{c}})$. Finally, when $\theta_X > \theta_C$, $\sqrt{k} L_{nk}^{-b} \left(\bar{E}_n - 1 \right)$ tends to 0 and D_n is thus the leading term : we obtain $\sqrt{k} L_{nk}^{1-b} \Delta_n \xrightarrow{d} N(0, \frac{a}{\tilde{c}})$.

The rest of the subsection is now devoted to the proof of Lemma 1 .
Let us introduce the function p defined by

$$p(x) = \mathbb{P}(\delta = 1 | Z = x).$$

Proceeding as in Einmahl et al. (2008), we carry on the proof by considering now that δ_i is related to Z_i by

$$\delta_i = \mathbb{I}_{U_i \leqslant p(Z_i)},$$

where $(U_i)_{i \leqslant n}$ denotes an independent sequence of standard uniform variables, independent of the sequence $(Z_i)_{i \leqslant n}$. We denote by $U_{[1,n]}, \ldots, U_{[n,n]}$ the (unordered) values of the uniform sample pertaining to the order statistics $Z_{1,n} \leqslant \ldots \leqslant Z_{n,n}$ of the observed sample $Z_1, \ldots, Z_n$.

Remind that $Z_i = \Lambda_H^-(E_i)$, where $E_1, \ldots, E_n$ are independent standard exponential random variables. We introduce, for every $1 \leqslant i \leqslant n$, the standard uniform random variables $V_i = 1 - \exp(-E_i)$ such that $Z_i = \Lambda_H^-(-\log(1 - V_i))$, and define the function

$$r(t) := (p \circ \Lambda_H^-)(-\log t).$$

Lemma 3 provides valuable information about the behavior of $r(\cdot)$ at infinity. We now write,

$$\begin{aligned}
D_n = \sqrt{k} L_{nk}^{-b} \left(L_{nk}^{1-a} \tfrac{\hat{p}_k}{\tilde{c}} - a \right) &= \frac{L_{nk}^{-b}}{\sqrt{k}} \sum_{j=1}^{k} \left(\frac{L_{nk}^{1-a}}{\tilde{c}} \mathbb{I}_{U_{[n-j+1,n]} \leqslant r(1 - V_{n-j+1,n})} - a \right) \\
&= \frac{L_{nk}^{b}}{\tilde{c}\sqrt{k}} \sum_{j=1}^{k} \left(\mathbb{I}_{U_{[n-j+1,n]} \leqslant r(1 - V_{n-j+1,n})} - \mathbb{I}_{U_{[n-j+1,n]} \leqslant r(j/n)} \right) \\
&\quad + \frac{L_{nk}^{-b}}{\sqrt{k}} \sum_{j=1}^{k} \left(\frac{L_{nk}^{1-a}}{\tilde{c}} \mathbb{I}_{U_{[n-j+1,n]} \leqslant r(j/n)} - a \right) \\
&=: T_{1,k} + T_{2,k}
\end{aligned}$$

Whatever the position of θ_X versus θ_C, we will prove below that the term $T_{1,k}$ above converges to 0 in probability. It turns out that this amounts to proving that, for some positive sequence $v_n = o(1/n)$ (to be

chosen later) and some constant $c > 0$,

$$\sqrt{k}L_{nk}^b S_{n,k} \stackrel{n\to\infty}{\longrightarrow} 0 \quad \text{where} \quad S_{n,k} := \sup\left\{ |r(s) - r(t)| \,;\, \frac{1}{n} \leqslant t \leqslant \frac{k}{n} \,,\, |s-t| \leqslant c\sqrt{k}/n \,,\, s \geqslant v_n \right\}. \qquad (18)$$

As a matter of fact, if we introduce the events

$$A_{n,c} = \left\{ \sup_{1\leqslant j\leqslant k} |(1 - V_{n-j+1,n}) - j/n| \leqslant c\sqrt{k}/n \right\} \quad \text{and} \quad B_n = \{1 - V_{n,n} \geqslant v_n\},$$

then, since $|\mathbb{I}_{U\leqslant a} - \mathbb{I}_{U\leqslant b}| \stackrel{d}{=} \mathbb{I}_{U\leqslant |a-b|}$ for any standard uniform U and constants a, b in $[0,1]$, it comes

$$\begin{aligned}
\mathbb{P}(|T_{1,k}| > \delta) &\leqslant \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} \mathbb{I}_{U_j \leqslant |r(1-V_{n-j+1,n}) - r(j/n)|} > \tilde{c}\delta/(\sqrt{k}L_{nk}^b) \right) \\
&\leqslant \mathbb{P}\left(\sqrt{k}L_{nk}^b S_{n,k} > \eta \right) + \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} \mathbb{I}_{U_j \leqslant \eta/(\sqrt{k}L_{nk}^b)} > \tilde{c}\delta/(\sqrt{k}L_{nk}^b) \right) + \mathbb{P}(B_n^c) + \mathbb{P}(A_{n,c}^c)
\end{aligned}$$

for any given $\delta > 0$ and $\eta > 0$. The second term in the right-hand side is (by Markov's inequality) lower than $\tilde{c}\delta/\eta$ (which is arbitrarily small), the third term is equal to $nv_n(1 + o(1)) = o(1)$, and the fourth term is arbitrarily small (for c large enough) by the weak convergence of the uniform tail quantile process. Therefore, we are left to prove that $\sqrt{k}L_{nk}^b S_{n,k} = o(1)$ (*i.e.* relation (18)), so that $T_{1,k} = o_{\mathbb{P}}(1)$ will be proved. This is done in the different cases distinguished below, along with the treatment of the main term $T_{2,k}$.

The whole proof heavily relies on the first and second order developments stated in Lemma 3 of the Appendix, concerning the function $p \circ \Lambda_H^-$.

1. Case $\theta_X < \theta_C$

In this situation, we have $a = 1$, $b = 0$, $\tilde{c} = 1$ and $p = \lim_{z\to+\infty} p(z) = \lim_{t\searrow 0} r(t) = 1$ (see Lemma 3). Hence

$$\begin{aligned}
T_{2,k} &= \tfrac{1}{\sqrt{k}}\textstyle\sum_{j=1}^{k} \left(\mathbb{I}_{U_{n-j+1,n} \leqslant r(j/n)} - 1 \right) \\
&\stackrel{d}{=} -\tfrac{1}{\sqrt{k}}\textstyle\sum_{j=1}^{k} \left(\mathbb{I}_{U_j > r(j/n)} - (1 - r(j/n)) \right) - \tfrac{1}{\sqrt{k}}\textstyle\sum_{j=1}^{k} (1 - r(j/n)) \\
&=: -T'_{2,k} - T''_{2,k},
\end{aligned}$$

where $T'_{2,k}$ turns out to be a sum of centered independent random variables. Let us now prove that $T'_{2,k} = o_{\mathbb{P}}(1)$, $T''_{2,k}$ tends to $A\alpha'$ (here $A = \frac{\theta_X}{\theta_C}\frac{c_G}{c_F^d}$ where α' is defined in condition $H_2(iii)$) and that $\sqrt{k}S_{n,k} \to 0$ (hence, as explained above, $T_{1,k} = o_{\mathbb{P}}(1)$).

Concerning $T'_{2,k}$, by definition of $r(\cdot)$ and thanks to Lemma 3 stated in the Appendix, we have

$$1 - r(x) = A(-\log x)^{d-1}(1 + o(x)) \quad \text{where} \quad d = \theta_X/\theta_C \in]0,1[.$$

Therefore, since $\log(n/j)/L_{nk}$ tends to 1 uniformly in j under condition H_1 (Lemma 5), we obtain

$$\mathbb{V}(T'_{2,k}) = \frac{1}{k}\sum_{j=1}^{k} r(j/n)(1 - r(j/n)) \leqslant \frac{1}{k}\sum_{j=1}^{k} (1 - r(j/n)) \leqslant L_{nk}^{d-1} A(1 + o(1)),$$

which implies that $\mathbb{V}(T'_{2,k})$ tends to 0, since $d < 1$.

Concerning $T''_{2,k}$, we have similarly, using now assumption $H_2(iii)$ and Lemma 5 ($\log n/j \sim L_{nk}$),

$$T''_{2,k} = A(1 + o(1))\sqrt{k}(L_{nk})^{d-1} \stackrel{n\to\infty}{\longrightarrow} A\alpha'.$$

Let us now deal with $\sqrt{k}S_{n,k}$. From now on, let *cst* denote some generic positive constant. Since $r(t)$ converges to 1 as $t \searrow 0$, and thanks to Lemma 3, we have, for s and t small,

$$\begin{aligned}
|r(s) - r(t)| &= \left| \frac{1}{r(s)} - \frac{1}{r(t)} \right| r(s)r(t) \\
&\leqslant cst \left\{ |(-\log t)^{d-1} - (-\log s)^{d-1}| + |(-\log t)^{d-1-\beta} v(-\log t) - (-\log s)^{d-1-\beta} v(-\log s)| \right\}
\end{aligned}$$

Introducing the set $Z_n = \{ (s,t) \,;\, 1/n \leqslant t \leqslant k/n \,,\, |t-s| \leqslant c\sqrt{k}/n \,,\, s \geqslant v_n \}$ and reminding that $v_n = o(1/n)$ (an appropriate sequence will be chosen in few lines), it can be checked that applying the mean value theorem to the function $h(t) = (-\log t)^{d-1}$ of positive derivative $h'(t) = (1-d)t^{-1}(-\log t)^{d-2}$, yields for large n

(below, $u = u(s,t)$ denotes some appropriate value between s and t)

$$\sqrt{k}\sup_{(s,t)\in Z_n}|h(t)-h(s)| \leqslant \sup_{(s,t)\in Z_n}|h'(u)|.|t-s| \leqslant cst\sqrt{k}\tfrac{1}{v_n}L_{nk}^{d-2}c\sqrt{k}/n = cst\tfrac{k}{nv_n}L_{nk}^{d-2}.$$

This is the first step towards the proof of $\sqrt{k}S_{n,k} = o(1)$. The second step requires to do the same job with the function $\tilde{h}(t) = (-\log t)^{d-1-\beta}v(-\log t)$, where $v(\cdot)$ is slowly varying at infinity. It is known (cf Bingham et. al. (1987) page 15) that we have $xv'(x)/v(x) \to 0$ and $x^{-\beta}v(x) \to 0$ as $x \to \infty$, so that

$$|\tilde{h}'(t)| = |1-d+\beta|\frac{1}{t}(-\log t)^{d-2}\left|1-cst\frac{xv'(x)}{v(x)}\right|x^{-\beta}|v(x)| \leqslant cst|h'(t)|$$

where x denoted $(-\log t)$, which is large when t is close to 0. Therefore, taking into account all the previous findings, and considering the choice $v_n = k^{-\epsilon}/n = o(1/n)$, we have proved that for n large

$$\sqrt{k}S_{n,k} \leqslant cst\tfrac{k}{nv_n}L_{nk}^{d-2} = cst.k^{1+\epsilon}L_{nk}^{d-2} = cst\left(\sqrt{k}L_{nk}^{(d-2)/2+\delta}\right)^{2(1+\epsilon)}$$

which turns out to be $o(1)$ as soon as $0 < \delta < d/2$ thanks to assumption $H_2(iii)$. This ends the proof of Lemma 1 in the mild censoring case $\theta_X < \theta_C$.

2. Case $\theta_X = \theta_C$

Here, we also have $a = 1$, $b = 0$ but now $\tilde{c} = \frac{c_F}{c_F+c_G} = p = \lim_{z\to\infty}p(z) = \lim_{t\searrow 0}r(t)$. It is clear that

$$\begin{aligned} T_{2,k} &\stackrel{d}{=} \frac{1}{p}\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\left(\mathbb{I}_{U_j\leqslant r(j/n)} - r(j/n)\right) + \frac{1}{p}\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(r(j/n)-p) \\ &=: T'_{2,k} + T''_{2,k} \end{aligned}$$

Let us prove that $T'_{2,k} \xrightarrow{d} N(0, \frac{1-p}{p})$, while $T''_{2,k}$ and $\sqrt{k}S_{n,k}$ are both $o(1)$.

Concerning $T'_{2,k}$: we have

$$\mathbb{V}(T'_{2,k}) = \frac{1}{p^2}\frac{1}{k}\sum_{j=1}^{k}r(j/n)(1-r(j/n)),$$

which tends to $\frac{1-p}{p}$, since $r(j/n)$ tends to p, uniformly in j (see Lemma 3). We conclude, for this term, using Lyapunov's theorem (details are omitted, here $r(j/n) \leqslant 1$).

Concerning $T''_{2,k}$, since Lemma 3 of the Appendix yields $r(t) = p\left(1-(-\log t)^{\rho}v(-\log t)\right)$, we have (for some $\delta > 0$)

$$T''_{2,k} \;=\; -\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(\log(n/j))^{\rho}v(\log(n/j)) \;=\; -\sqrt{k}(L_{nk})^{\rho+\delta}L_{nk}^{-\delta}v(L_{nk})\frac{1}{k}\sum_{j=1}^{k}u_{n,j}^{\rho}$$

where we noted $u_{n,j} = \log(n/j)/L_{nk}$, which tends to 1 uniformly in j thanks to condition H_1, and used the fact that $v(\log(n/j)) \sim v(L_{nk})$ because $v \in RV_0$. The Riemann sum on the right-hand side converges to 1, so for a choice of δ satisfying assumption $H_3(i)$, we have proved that $T''_{2,k} = o(1)$.

Concerning now $\sqrt{k}S_{n,k}$, we proceed similarly as in the first case. Introducing $\tilde{h}(t) = (-\log t)^{\rho}v(-\log t)$ where $v(\cdot)$ is slowly varying at infinity, we have as previously $|\tilde{h}'(t)| = \frac{1}{t}(-\log t)^{\rho-1+\epsilon}o(1)$ for $t \searrow 0$ and any some small $\epsilon > 0$. Therefore, Lemma 3, definitions of $S_{n,k}$ and of the set Z_n, along with the mean value theorem, yield

$$\sqrt{k}S_{n,k} = \tilde{c}\sup_{(s,t)\in Z_n}|\tilde{h}(t)-\tilde{h}(s)| \leqslant cst\sqrt{k}\sup_{(s,t)\in Z_n}\{|\tilde{h}'(u)|.|t-s|\} \leqslant cst\sqrt{k}\frac{1}{v_n}L_{nk}^{\rho-1+\epsilon}\tilde{c}\frac{\sqrt{k}}{n}.$$

Choosing, in the definition of $S_{n,k}$, the sequence $v_n = k^{-\epsilon}/n = o(1/n)$ for some small $\epsilon > 0$, we have

$$\sqrt{k}S_{n,k} = cst\left(\sqrt{k}L_{nk}^{(\rho-1+\epsilon)/(2(1+\epsilon))}\right)^{2(1+\epsilon)} = cst\left(\sqrt{k}L_{nk}^{(\rho-1)/2+\delta}\right)^{2(1+\epsilon)}$$

which turns out to be $o(1)$ according to assumption $H_3(i)$ (if $\rho \geqslant 1$) or $H_3(ii)$ (if $\rho < 0$), as soon as δ is sufficiently small. This ends the proof of Lemma 1 in the semi-strong censoring case $\theta_X = \theta_C$.

3. Case $\theta_X > \theta_C$

Now we are in the situation where $a < 1$, $b = (1-a)/2 \in]0, 1/2[$, $\tilde{c} = \frac{c_F}{c_G^a}$ and $p = \lim_{z\to\infty}p(z) = \lim_{t\searrow 0}r(t) =$

0. Since $1-a-b=b$, we have readily

$$\begin{aligned} T_{2,k} &\stackrel{d}{=} \frac{L_{nk}^b}{\tilde{c}}\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\left(\mathbb{I}_{U_j\leqslant r(j/n)}-r(j/n)\right)+\frac{aL_{nk}^{-b}}{\sqrt{k}}\sum_{j=1}^{k}\left(\frac{L_{nk}^{1-a}}{a\tilde{c}}r(j/n)-1\right)\\ &=: T'_{2,k}+T''_{2,k} \end{aligned}$$

Let us prove that $T'_{2,k}\xrightarrow{d}N(0,\frac{a}{\tilde{c}})$, while $T''_{2,k}$ and $\sqrt{k}L_{nk}^b S_{n,k}$ are both $o(1)$ (the latter will guarantee that $T_{1,k}=o_{\mathbb{P}}(1)$).

Concerning $T'_{2,k}$: we have

$$\mathbb{V}(T'_{2,k})=\frac{L_{nk}^{2b}}{\tilde{c}^2}\frac{1}{k}\sum_{j=1}^{k}r(j/n)(1-r(j/n))$$

Lemma 3 in the Appendix yields the following first order development, as $t\searrow 0$,

$$r(t)=a\tilde{c}(-\log t)^{a-1}(1+o(t))=a\tilde{c}(-\log t)^{-2b}(1+o(t)). \tag{19}$$

Since $u_{n,j}=\log(n/j)/L_{nk}$ tends to 1 uniformly in j, under condition H_1 (see Lemma 5), it is then easy to see that $\mathbb{V}(T'_{2,k})$ tends to $\frac{a}{\tilde{c}}$. We conclude concerning $T'_{2,k}$ using Lyapunov's theorem (again, details are easy and omitted).

Concerning $T''_{2,k}$: we write

$$\frac{L_{nk}^{1-a}}{a\tilde{c}}r(j/n)-1=\left(\frac{L_{nk}^{1-a}}{a\tilde{c}}r(j/n)-\left(\frac{L_{nk}}{\log(n/j)}\right)^{1-a}\right)+\left(\left(\frac{L_{nk}}{\log(n/j)}\right)^{1-a}-1\right)$$

and treat these two terms separately. Using the second order formula stated in Lemma 3, we have

$$\frac{1}{r(t)}=1+\frac{(-\log t)^{1-a}}{a\tilde{c}}\left(1-(-\log t)^{\tilde{\rho}}v(-\log t)\right). \tag{20}$$

and consequently, for some small $\delta>0$,

$$\begin{aligned} \frac{a\tilde{c}}{L_{nk}^{1-a}r(j/n)} &= \left(\frac{\log(n/j)}{L_{nk}}\right)^{1-a}\left(1-(\log(n/j))^{\tilde{\rho}}v(\log(n/j))+a\tilde{c}(\log(n/j))^{a-1}\right)\\ &= \left(\frac{\log(n/j)}{L_{nk}}\right)^{1-a}\left(1-L_{nk}^{\tilde{\rho}+\delta}o(1)+a\tilde{c}L_{nk}^{a-1}(1+o(1))\right) \end{aligned}$$

where we used condition H_1 and the slow variation of v, which guarantees that $v(\log(n/j))\sim v(L_{nk})$ and $x^{-\delta}v(x)\to 0$ as $x\to\infty$. Now, since $\tilde{\rho}=\max(\theta_Z\rho_F,\theta_Z\rho_G,a-1)\geqslant a-1$, it comes

$$\frac{L_{nk}^{1-a}}{a\tilde{c}}r(j/n)-\left(\frac{L_{nk}}{\log(n/j)}\right)^{1-a}=(1+o(1))L_{nk}^{\tilde{\rho}+\delta}o(1)$$

and therefore the first term of $T''_{2,k}$ is equal to $a\sqrt{k}L_{nk}^{-b+\tilde{\rho}+\delta}o(1)$, which tends to 0 under condition $H_4(ii)$. The second term of $T''_{2,k}$ is

$$a\sqrt{k}L_{nk}^{-b}\frac{1}{k}\sum_{j=1}^{k}\left(\left(\frac{L_{nk}}{\log(n/j)}\right)^{1-a}-1\right).$$

But $\left(\frac{L_{nk}}{\log(n/j)}\right)^{1-a}-1=(a-1)\frac{\log(k/j)}{L_{nk}}(1+o(1))$ with $\frac{1}{k}\sum_{j=1}^{k}\log(k/j)$ tending to 1. So the second term of $T''_{2,k}$ is equal to

$$a(a-1)\sqrt{k}L_{nk}^{-1-b}(1+o(1)),$$

and this quantity tends to 0 under condition $H_4(iv)$.

Concerning now $\sqrt{k}L_{nk}^b S_{n,k}$, we have

$$S_{n,k}=\sup_{(s,t)\in Z_n}|r(t)-r(s)|\leqslant\sup_{(s,t)\in Z_n}\left|\frac{1}{r(t)}-\frac{1}{r(s)}\right|\sup_{(s,t)\in Z_n}\{r(t)r(s)\}.$$

Thanks to the first order relation (19), the second supremum of the right-hand side is lower than a constant times $L_{nk}^{2(a-1)}$. The first supremum will be handled with the more precise second order development (20),

which yields

$$\sup_{(s,t)\in Z_n}\left|\frac{1}{r(t)}-\frac{1}{r(s)}\right| = cst\left\{\sup_{(s,t)\in Z_n}|h(t)-h(s)|+\sup_{(s,t)\in Z_n}|\tilde{h}(t)-\tilde{h}(s)|\right\}$$

where we define $h(t)=(-\log t)^{1-a}$ and $\tilde{h}(t)=(-\log t)^{1-a+\tilde{\rho}}v(-\log t)$. Contrary to the functions arisen in case 1, the functions h and $\tilde{h}$ tend to infinity instead of vanishing to 0, when $t\searrow 0$: this will be counterbalanced by the second supremum. Studying derivatives of the functions h and $\tilde{h}$, and again using a first order Taylor expansion, we obtain via similar computations as in the previous cases, for n large and any $\epsilon>0$ (with the choice $v_n=k^{-\epsilon}/n$),

$$\sup_{(s,t)\in Z_n}\left|\frac{1}{r(t)}-\frac{1}{r(s)}\right|\leqslant cst.k^{1/2+\epsilon}L_{nk}^{-a}.$$

Therefore, gathering the two suprema, we have (for some small value of $\delta>0$ depending on ϵ)

$$\sqrt{k}L_{nk}^{b}S_{n,k}\leqslant cst.k^{1+\epsilon}L_{nk}^{b-a}L_{nk}^{2(a-1)}=cst.k^{1+\epsilon}L_{nk}^{-1-b}=cst\left(\sqrt{k}L_{nk}^{-(1+b)/2+\delta}\right)^{2(1+\epsilon)}$$

which, by assumption $H_4(iii)$, converges to 0 as $n\to\infty$.

3.2. Proof of Proposition 2

Remind from (14) that

$$R_{n,l}=\frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{l(E_{n-j+1,n})}{l(E_{n-k,n})}\right)\text{ and }R_{n,\tilde{l}}=\frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})}\right).$$

Let $A>1$. Under condition $R_l(B,\rho)$, we have for all $\epsilon>0$ and t sufficiently large

$$(1-\epsilon)B(t)K_\rho(x)\leqslant\frac{l(tx)}{l(t)}-1\leqslant(1+\epsilon)B(t)K_\rho(x)\quad(\forall 1\leqslant x\leqslant A).$$

We only prove the result for $R_{n,l}$, the proof for $R_{n,\tilde{l}}$ being very similar, using $R_{\tilde{l}}(\tilde{B},\tilde{\rho})$ instead of $R_l(B,\rho)$. Note that

$$R_{n,l}=\frac{1}{k}\sum_{j=1}^{k}\log\left(1+\xi_{j,n}\right),$$

where $\xi_{j,n}=\frac{l(E_{n-j+1,n})}{l(E_{n-k,n})}-1$ tends to 1 uniformly in j, because l is slowly varying and $\frac{E_{n-j+1,n}}{E_{n-k,n}}$ tends to 1 uniformly in j, according to Lemma 5. Hence, using the following inequality,

$$x-x^2/2\leqslant\log(1+x)\leqslant x\quad(\forall x\geqslant -1/2)$$

and the fact that $x_{j,n}:=\frac{E_{n-j+1,n}}{E_{n-k,n}}\geqslant 1$ tends to 1 uniformly in j, we obtain that for all $\epsilon>0$ and n sufficiently large,

$$R_{n,l}\leqslant\frac{1}{k}\sum_{j=1}^{k}\left(\frac{l(E_{n-j+1,n})}{l(E_{n-k,n})}-1\right)\leqslant(1+\epsilon)B(E_{n-k,n})\frac{1}{k}\sum_{j=1}^{k}K_\rho(x_{j,n}),$$

omitting the lower bound, which is treated similarly. Since $K_\rho(1+x)\sim x$ when x tends to 0, then $K_\rho(x_{j,n})\sim\frac{E_{n-j+1,n}-E_{n-k,n}}{E_{n-k,n}}$, uniformly in j. By Lemma 4, $\frac{E_{n-j+1,n}-E_{n-k,n}}{E_{n-k,n}}\overset{d}{=}\frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}}$. Hence, it is easy to prove that

$$E_{n-k,n}\frac{1}{k}\sum_{j=1}^{k}K_\rho(x_{j,n})\xrightarrow{\mathbb{P}}1.$$

Since B is regularly varying and $\frac{E_{n-k,n}}{L_{nk}}\to 1$, then $\frac{B(E_{n-k,n})}{E_{n-k,n}}\sim\frac{B(L_{nk})}{L_{nk}}$ and consequently

$$\sqrt{k}L_{nk}^{-b}B(L_{nk})(1+o_{\mathbb{P}}(1))\leqslant\liminf\sqrt{k}L_{nk}^{1-b}R_{n,l}\leqslant\limsup\sqrt{k}L_{nk}^{1-b}R_{n,l}\leqslant\sqrt{k}L_{nk}^{-b}B(L_{nk})(1+o_{\mathbb{P}}(1)).$$

We conclude using assumption $R_l(B,\rho)$ and conditions $H_2(i)$, $H_3(i)$ or $H_4(ii)$, because $|B|$ is regularly varying of order ρ, and we have $\rho=\tilde{\rho}$ when $\theta_X\leqslant\theta_C$, and $\rho\leqslant\tilde{\rho}$ when $\theta_X>\theta_C$ (see Lemma 2).

3.3. Proof of Proposition 3

Recall that

$$M_n = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{E_{n-j+1,n}}{E_{n-k,n}}\right).$$

Since $\frac{E_{n-j+1,n}}{\log(n/j)} \xrightarrow{\mathbb{P}} 1$ and $\frac{L_{nk}}{\log(n/j)} \xrightarrow{\mathbb{P}} 1$, uniformly in $j = 1,\ldots,k$ (see Lemma 5), then $\frac{E_{n-j+1,n}}{E_{n-k,n}} \xrightarrow{\mathbb{P}} 1$, uniformly in $j = 1,\ldots,k$. By Lemma 4, $(E_{n-j+1,n} - E_{n-k,n})_{1\leqslant j\leqslant k} \stackrel{d}{=} (\tilde{E}_{k,k},\ldots,\tilde{E}_{1,k})$. Therefore

$$M_n \stackrel{d}{=} \frac{1}{k}\sum_{j=1}^{k} \log\left(1 + \frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}}\right) = (1+o_{\mathbb{P}}(1))\frac{1}{E_{n-k,n}}\frac{1}{k}\sum_{j=1}^{k}\tilde{E}_j,$$

with $\frac{1}{k}\sum_{j=1}^{k}\tilde{E}_j \to 1$, a.s. Hence, $L_{nk}M_n$ also tends to 1, in probability.

4. Proof of Theorem 2

Starting from $x_{p_n} = \bar{F}^{-1}(p_n)$ and the definition of $\hat{x}_{p_n}$ in (9), we obtain

$$\begin{aligned}
\log(x_{p_n}) &= \theta_X \log\log(1/p_n) + \log(\bar{l}_F(-\log(p_n))),\\
\log(\hat{x}_{p_n}) &= \hat{\theta}_{X,k}\log\log(1/p_n) - \hat{\theta}_{X,k}\log(\hat{\Lambda}_{nF}(Z_{n-k,n})) + \log(Z_{n-k,n}).
\end{aligned}$$

Hence

$$\begin{aligned}
\log(\hat{x}_{p_n}/x_{p_n}) &= (\hat{\theta}_{X,k}-\theta_X)\log\log(1/p_n) - \hat{\theta}_{X,k}\log\left(\tfrac{\hat{\Lambda}_{nF}}{\Lambda_F}(Z_{n-k,n})\right) - (\hat{\theta}_{X,k}-\theta_X)\log(\Lambda_F(Z_{n-k,n}))\\
&\qquad + \left\{-\log(\bar{l}_F(\log(1/p_n))) - \theta_X\log(l_F(Z_{n-k,n}))\right\},\\
&=: Q_{1,n} + Q_{2,n} + Q_{3,n} + Q_{4,n}.
\end{aligned}$$

First of all, the result of Theorem 1 implies that

$$\frac{\sqrt{k}L_{nk}^{-b}}{\log\log(1/p_n)}Q_{1,n} = \sqrt{k}L_{nk}^{-b}(\hat{\theta}_{X,k}-\theta_X) \xrightarrow{d} N\left(m, \frac{\theta_X^2}{a\tilde{c}}\right).$$

Then, Lemma 6 (stated in the Appendix) implies that $(\hat{\Lambda}_{nF}/\Lambda_F)(Z_{n-k,n}) - 1 = O_{\mathbb{P}}\left(1/(\sqrt{k}\Lambda_F(Z_{n-k,n}))\right)$. Hence

$$\frac{\sqrt{k}L_{nk}^{-b}}{\log\log(1/p_n)}Q_{2,n} = O_{\mathbb{P}}(1)\frac{1}{L_{nk}^{b}\log\log(1/p_n)\Lambda_F(Z_{n-k,n})} \xrightarrow{\mathbb{P}} 0.$$

Now, remind that $\Lambda_F(Z_{n-k,n}) = \Lambda_F \circ \Lambda_H^{-}(E_{n-k,n}) = E_{n-k,n}^{a}\tilde{l}(E_{n-k,n})$. Hence, the asymptotic normality of $(\hat{\theta}_{X,k}-\theta_X)$ yields

$$\frac{\sqrt{k}L_{nk}^{-b}}{\log\log(1/p_n)}Q_{3,n} = O_{\mathbb{P}}(1)\frac{\log(L_{nk})}{\log\log(1/p_n)}\left(a\frac{\log(E_{n-k,n})}{\log(L_{nk})} + \frac{\log(\tilde{l}(E_{n-k,n}))}{\log(L_{nk})}\right).$$

The additional condition H_1' of Theorem 2, along with Lemma 5, imply that this term tends to 0 in probability.

Finally, Lemma 2 implies that

$$Q_{4,n} = -\log\left(1 - \log(1/p_n)^{\theta_X\rho_F}\bar{v}(\log(1/p_n))\right) - \theta_X\log\left(1 - Z_{n-k,n}^{\rho_F}v(Z_{n-k,n})\right),$$

where v and $\bar{v}$ are slowly varying. Hence, $\frac{\sqrt{k}L_{nk}^{-b}}{\log\log(1/p_n)}Q_{4,n}$ tends to 0 as soon as there exist some $0<\delta<1$ such that $\frac{\sqrt{k}L_{nk}^{-b}}{\log\log(1/p_n)}(\log 1/p_n)^{\theta_X\rho_F+\delta} = O(1)$ and $\frac{\sqrt{k}L_{nk}^{-b}}{\log\log(1/p_n)}Z_{n-k,n}^{\rho_F+\delta} = O_{\mathbb{P}}(1)$. Remind that $Z_{n-k,n} = E_{n-k,n}^{\theta_Z}l(E_{n-k,n})$. Hence, condition H_1' guarantees that we only need to show that $\sqrt{k}L_{nk}^{-b+\theta_X\rho_F} = O(1)$ and $\sqrt{k}L_{nk}^{-b+\theta_Z\rho_F} = O(1)$. When $\theta_X = \theta_Z < \theta_C$, this is due to the additional condition $H_2(iv)$. When $\theta_X = \theta_Z = \theta_C$, it is due to condition $H_3(i)$. Finally, when $\theta_X > \theta_Z = \theta_C$, it is due to $H_4(ii)$.

5. Finite sample comparisons

In this section, we illustrate, using a few simulations, the finite sample performances of our estimators of θ_X and x_{p_n} (for small p_n), in terms of observed bias and mean squared error (MSE). The sample size $n = 500$ has been considered.

We consider two classes of distributions of Weibull-tail type, for the target X and the censoring variable C :

- Weibull$(1/\theta)$ with c.d.f. $1 - \exp(-x^{1/\theta})$ $(x > 0)$, which Weibull-tail coefficient is θ.
- Gamma(a, b) with c.d.f. $\int_0^x \Gamma(a)^{-1} b^a u^{a-1} e^{-bu} du$ $(x > 0)$, which Weibull-tail coefficient is 1.

We consider two cases : a Weibull distribution censored by another Weibull distribution and a Gamma distribution censored by a Weibull distribution. In each case, we consider three situations with $\theta_X < \theta_C$, $\theta_X = \theta_C$ or $\theta_X > \theta_C$, corresponding to different intensities of the censoring in the tail.

In Figures 1 and 2, we compare our estimator $\hat{\theta}_{X,k}$ defined in (7) with the estimator $\hat{\theta}_X^{(complete)}$ defined in (5), which is applied to the X sample as if it was observed (of course, in practice, it is not, so the comparison is of theoretical interest only). We also compare it with $\hat{\theta}_{Z,k}$ defined in (8), which is the same expression as $\hat{\theta}_X^{(complete)}$ but applied to the observed sample Z.
For each considered distribution, 2000 random samples of length $n = 500$ are generated ; median bias and MSE of the above-mentioned estimators are plotted against different values of k_n, the number of excesses used.

This small simulation study shows that for finite sample sizes, as expected, using an estimator (here $\hat{\theta}_{Z,k}$) not adapted to censoring yields inaccurate results, even in the case $\theta_X < \theta_C$, where $\hat{\theta}_{Z,k}$ is consistent for estimating θ_X. The case $\theta_X = \theta_C$ is particular : since the censored and the censoring distributions have the same Weibull-tail coefficient, it seems that $\hat{\theta}_{Z,k}$ succeeds in reaching its target. We also see that our proposed estimator $\hat{\theta}_{X,k}$ has good performances, comparable to those of $\hat{\theta}_X^{(complete)}$ in the non-censored case. Note that the bias of $\hat{\theta}_{X,k}$ does not vary much, whereas the MSE deteriorates when the censoring becomes stronger.

In Figure 3, we illustrate the result of Remark 4 concerning the convergence of $\hat{p}_k$, the proportion of non-censored data in the tail, when the sample size n tends to $+\infty$. For three different situations corresponding to $\theta_X < \theta_C$ ($\hat{p}_k \to p = 1$), $\theta_X = \theta_C$ ($\hat{p}_k \to p = 1/2$ for the distributions considered) or $\theta_X > \theta_C$ ($\hat{p}_k \to p = 0$), we present plots of $\hat{p}_k$ against k_n for three sample sizes $n = 500, 5000, 50000$.

We can see that the convergence of $\hat{p}_k$ is very slow : In particular, in practice, when $\theta_X < \theta_C$, the proportion of non-censored data in the tail is quite far from 1. This explains why one needs to take the censoring mechanism into account for the estimation of θ_X and why the existing estimators, defined in the non-censored setting, should not be used.

Now in Figures 4 and 5, for the value $p_n = 0.01$, we compare our estimator $\hat{x}_{p_n}$ defined in (9) with the following estimator

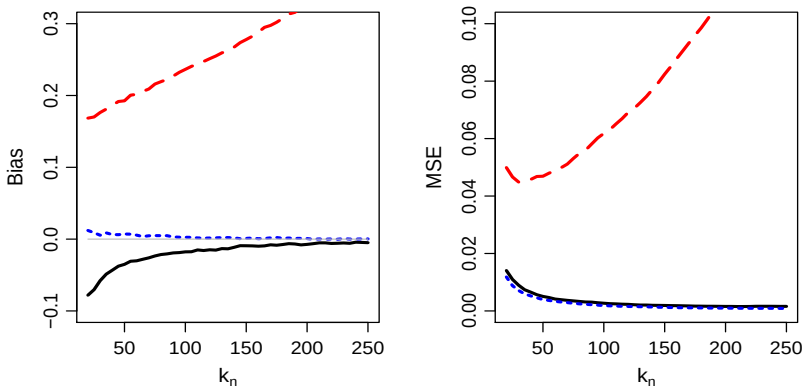
$$\hat{x}_{Z,p_n} = Z_{n-k,n} \left(\frac{-\log p_n}{\log(n/k_n)} \right)^{\hat{\theta}_{Z,k}},$$

which is the estimator defined in the non-censored setting, in Gardes and Girard (2005), but applied to the observed sample Z. We also compare it with the existing estimator defined, in a more general censored setting, by equation (8) in Einmahl et al. (2008) :

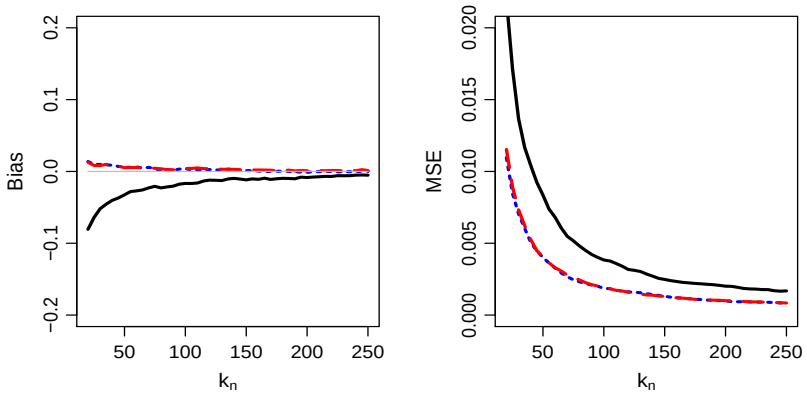
$$\hat{x}_{p_n}^{EFG} = Z_{n-k,n} + \hat{a}_k \frac{((1 - \hat{F}_n(Z_{n-k}))/p_n)^{\hat{\gamma}^{c,Mom}} - 1}{\hat{\gamma}^{c,Mom}},$$

where $\hat{\gamma}^{c,Mom}$ is the moment estimator of the extreme value index γ_X of F adapted to censoring (note that γ_X is equal to 0 in the Weibull-tail situation), and $\hat{F}_n$ stands for the Kaplan-Meier estimator of the c.d.f. F. We refer to Einmahl et al. (2008) for the expression of $\hat{a}_k$.

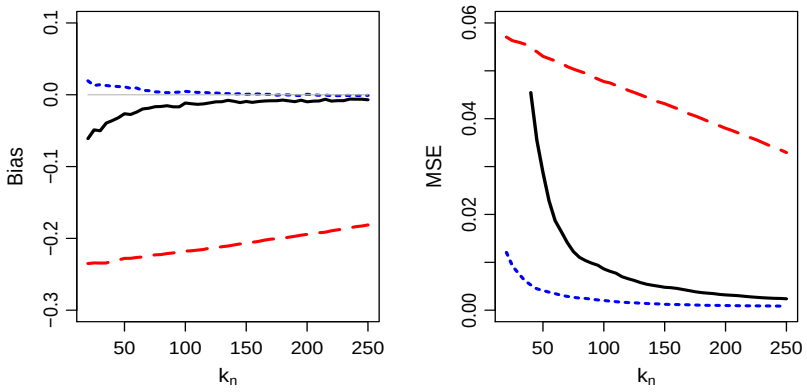
We clearly see that the estimator $\hat{x}_{Z,p_n}$ (not adapted to censoring) is inaccurate for the estimation of x_{p_n}. Indeed, $\hat{x}_{Z,p_n}$ actually estimates the quantile of $\bar{H} = \bar{F}\bar{G} \leqslant \bar{F}$ and therefore underestimates x_{p_n}, as showed in our plots. We also show that our proposed estimator is competitive, especially in term of bias, although the MSE is often higher than that of the existing estimator $\hat{x}_{p_n}^{EFG}$.



(a) Weibull(2) censored by Weibull(2/3), where $\theta_X = 1/2$ and $\theta_C = 3/2$.

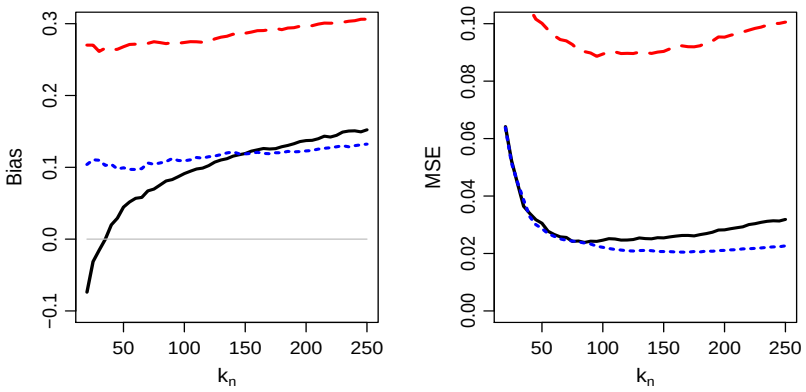


(b) Weibull(2) censored by Weibull(2), where $\theta_X = 1/2 = \theta_C$.

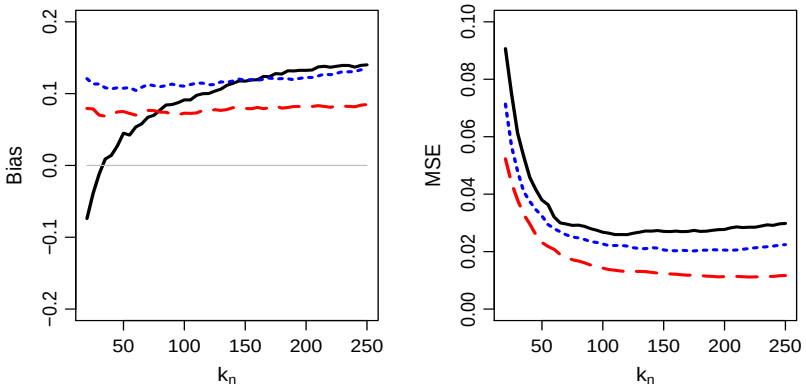


(c) Weibull(2) censored by Weibull(5), where $\theta_X = 1/2$ and $\theta_C = 1/5$.

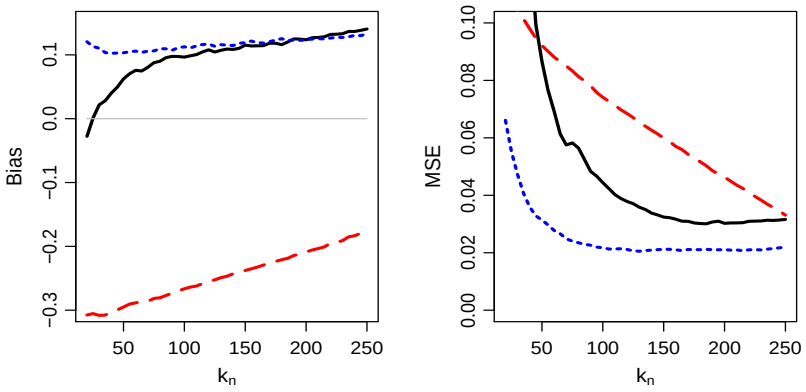
Figure 1: Comparison of bias and MSE of the estimators $\hat{\theta}_{X,k}$ (solid black), $\hat{\theta}_X^{(complete)}$ (dotted blue) and $\hat{\theta}_{Z,k}$ (dashed red) for a Weibull distribution censored by another Weibull distribution.



(a) Gamma$(0.8, 1)$ censored by Weibull$(2/3)$, where $\theta_X = 1$ and $\theta_C = 3/2$.



(b) Gamma$(0.8, 1)$ censored by Weibull(1), where $\theta_X = 1 = \theta_C$.



(c) Gamma$(0.8, 1)$ censored by Weibull(2), where $\theta_X = 1$ and $\theta_C = 1/2$.

Figure 2: Comparison of bias and MSE of the estimators $\hat{\theta}_{X,k}$ (solid black), $\hat{\theta}_X^{(complete)}$ (dotted blue) and $\hat{\theta}_{Z,k}$ (dashed red) for a Gamma distribution censored by a Weibull distribution.

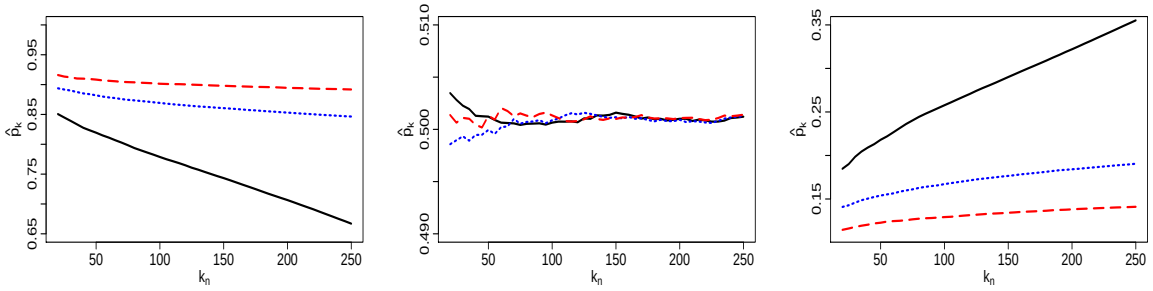
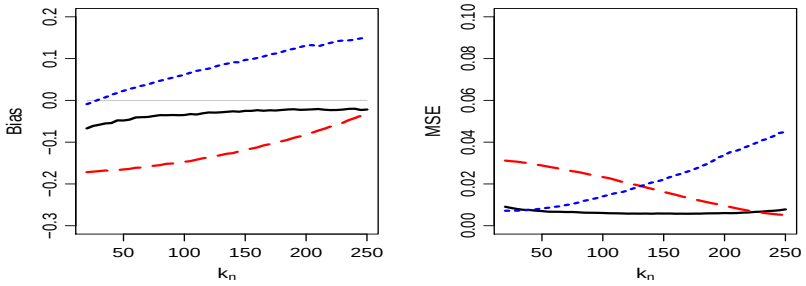
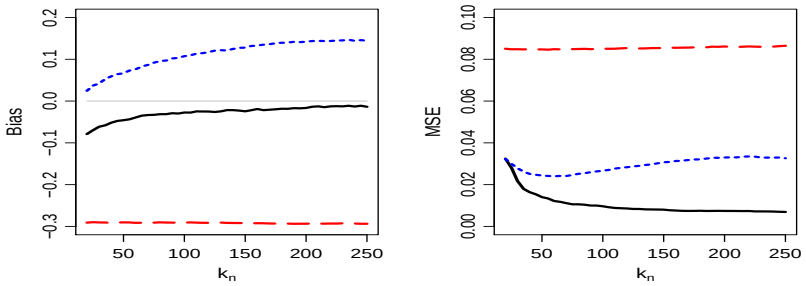


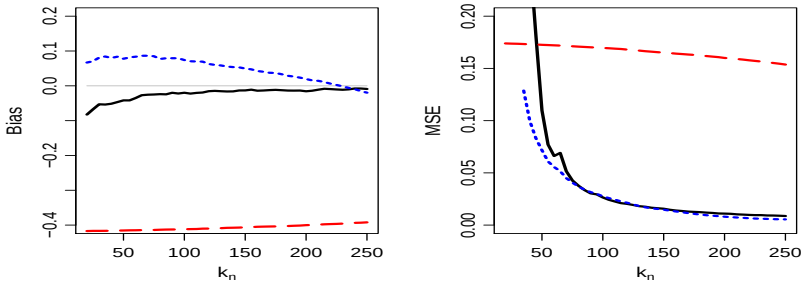
Figure 3: Comparison of the proportion of non-censoring in the tail $\hat{p}_k$ for Weibull(2) ($\theta_X = 1/2$) censored, respectively, by Weibull(2/3) ($\theta_C = 3/2 > \theta_X$, $p = 1$), Weibull(2) ($\theta_C = 1/2 = \theta_X$, $p = 1/2$) and Weibull(5) ($\theta_C = 1/5 < \theta_X$, $p = 0$) for $n = 500$ (solid black), $n = 5000$ (dotted blue) and $n = 50000$ (dashed red).



(a) Weibull(2) censored by Weibull(2/3), where $\theta_X = 1/2$ and $\theta_C = 3/2$.

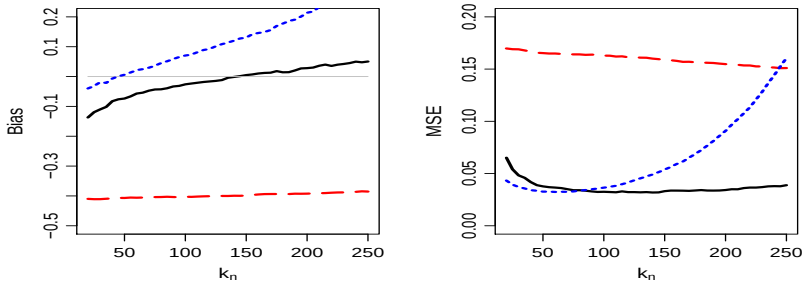


(b) Weibull(2) censored by Weibull(2), where $\theta_X = 1/2 = \theta_C$.

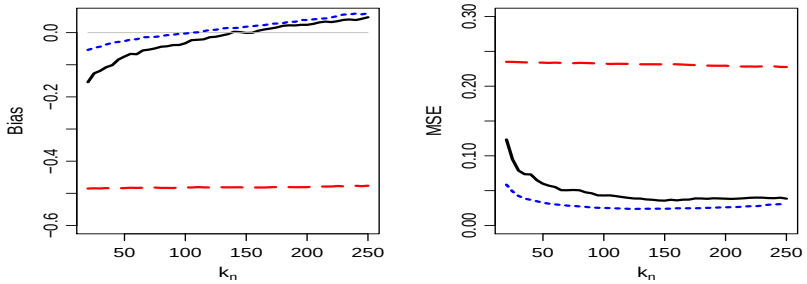


(c) Weibull(2) censored by Weibull(5), where $\theta_X = 1/2$ and $\theta_C = 1/5$.

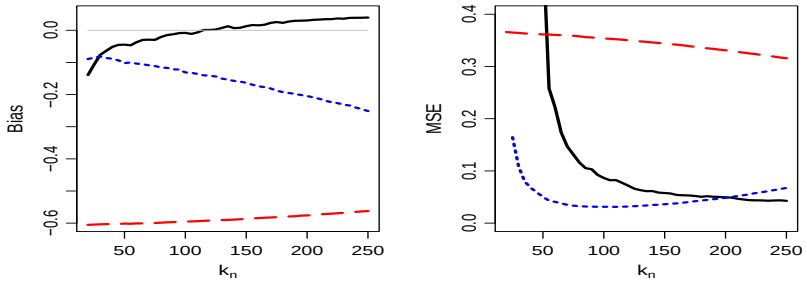
Figure 4: Comparison of bias and MSE of the estimators $\hat{x}_{p_n}$ (solid black), $\hat{x}^Z_{p_n}$ (dashed red) and $\hat{x}^{EFG}_{p_n}$ (dotted blue) of the quantile x_{p_n} ($p_n = 0.01$), for a Weibull distribution censored by another Weibull distribution.



(a) Gamma$(0.8, 1)$ censored by Weibull$(2/3)$, where $\theta_X = 1$ and $\theta_C = 3/2$.



(b) Gamma$(0.8, 1)$ censored by Weibull(1), where $\theta_X = 1 = \theta_C$.



(c) Gamma$(0.8, 1)$ censored by Weibull(2), where $\theta_X = 1$ and $\theta_C = 1/2$.

Figure 5: Comparison of bias and MSE of the estimators $\hat{x}_{p_n}$ (solid black), $\hat{x}^Z_{p_n}$ (dashed red) and $\hat{x}^{EFG}_{p_n}$ (dotted blue) of the quantile x_{p_n} $(p_n = 0.01)$, for a Gamma distribution censored by a Weibull distribution.

References

B. Brahimi, D. Meraghni and A. Necir Approximations to the tail index estimator of a heavy-tailed distribution under random censoring and application. In *Mathematical Methods in Statistic.***24**, pages 266-279 (2015)

B. Brahimi, D. Meraghni and A. Necir . Nelson-Aalen tail product-limit process and extreme value index estimation under random censorship. Unpublished manuscript, available on the ArXiv archive : https://arxiv.org/abs/1502.03955v2 (2016)

B. Brahimi, D. Meraghni, A. Necir and L. Soltane. Tail empirical process and a weighted extreme value index estimator for randomly right-censored data. Unpublished manuscript, available on the ArXiv archive : https://arxiv.org/abs/1801.00572 (2018)

J. Beirlant, G. Dierckx, A. Guillou and A. Fils-Villetard . Estimation of the extreme value index and extreme quantiles under random censoring. In *Extremes* **10**, pages 151-174 (2007)

J. Beirlant, M. Broniatowski, J. Teugels and P. Vynckier. The mean residual life function at great age : Applications to tail estimation. In *Journal of Statistical Planning and Inference* **45**, pages 21-48 (1995)

J. Beirlant, Y. Goegebeur, J. Segers and J. Teugels Statistics of extremes: Theory and applications Wiley (2004)

J. Beirlant, A. Bardoutsos, T. de Wet and I. Gijbels . Bias reduced tail estimation for censored Pareto type distributions. In *Stat. Prob. Letters* **109**, pages 78-88 (2016)

J. Beirlant, G. Maribe and A. Verster . Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications . In *Insurance : Mathematics and Economics* **78**, pages 114-122 (2018)

J. Beirlant, J. Worms and R. Worms. Asymptotic distribution for an extreme value index estimator in a censorship framework. To appear in *Journal of Statistical Planning and Inference* (2019).

N. H. Bingham, C.M. Goldie and J.L. Teugels (1987) Regular variation. In *Cambridge University Press.*

S. Csorgo. Probability theory. Independence, interchangeability, martingales. In *The Annals of Statistics* **24(6)**, pages 2744-2778 (1996)

L. de Haan and A. Ferreira . Extreme Value Theory : an Introduction. *Springer Science + Business Media* (2006)

J. Diebolt, L. Gardes, S. Girard and A. Guillou Bias-reduced estimators of the Weibull tail-coefficient. In *Test* **17**, pages 311-331 (2008)

G. Dierckx, J. Beirlant, D. de Waal and A.Guillou A new estimation method for Weibull-type tails based on the mean excess function. In *Journal of Statistical Planning and Inference* **139**, pages 1905-1920 (2009)

A. Diop, J-F. Dupuy and P. Ndao. Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. In *Computational Statistics & Data Analysis* **79**, pages 63-79 (2014)

J. Einmahl, A. Fils-Villetard and A. Guillou . Statistics of extremes under random censoring. In *Bernoulli* **14**, pages 207-227 (2008)

L. Gardes and S. Girard . Estimating extreme quantiles of Weibull-Tail distributions. In *Communications in Statistics : Theory and Methods* **34**, pages 1065-1080 (2005)

S. Girard . A Hill type estimator of the Weibull-tail coefficient. In *Communications in Statistics : Theory and Methods* **33(2)**, pages 205-234 (2004)(a)

S. Girard . A Hill type estimator of the Weibull-tail coefficient. HAL archive version : hal-00724602 (2004)(b)

Y. Goegebeur, J. Beirlant and T. de Wet Generalized Kernel Estimators for the Weibull-Tail Coefficient. In *Communications in Statistics : Theory and Methods* **39**, pages 3695-3716 (2010)

M.I. Gomes and M.M. Neves (2011). . Estimation of the extreme value index for randomly censored data. In *Biometrical Letters* **48 (1)**, pages 1-22.

P. Ndao., A. Diop, and J-F. Dupuy Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. In *Computational Statistics & Data Analysis* **79**, pages 63-79 (2014)

P. Ndao., A. Diop, and J-F. Dupuy Nonparametric estimation of the conditional extreme-value index with random covariates and censoring. In *Journal of Statistical Planning and Inference,* **168**, pages 20-37 (2016)

R.D Reiss. Approximate distribution of order statistics. *Springer-Verlag* (1989)

G. Stupfler . Estimating the conditional extreme-value index in presence of random right-censoring. In *Journal of Multivariate Analysis* **144**, pages 1-24 (2016)

J. Worms and R. Worms. New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. In *Extremes* **17 (2)**, pages 337-358 (2014)

J. Worms and R. Worms. Moment estimators of the extreme value index for randomly censored data in the Weibull domain of attraction. Unpublished manuscript, available on the ArXiv archive, arXiv:1506.03765 (2015).

J. Worms and R. Worms. Extreme value statistics for censored data with heavy tails under competing risks. In *Metrika* **81 (7)**, pages 849-889 (2018).

M. Zhou (1991). Some Properties of the Kaplan-Meier Estimator for Independent Non identically Distributed Random Variables. In *Annals of Statistics* **19 (4)**, pages 2266-2274 (1991)

6. Appendix

6.1. Details on the second order conditions

Remind that the starting assumption of this paper is relation (3),

$$\Lambda_F(x) = x^{1/\theta_X} l_F(x) \text{ and } \Lambda_G(x) = x^{1/\theta_C} l_G(x),$$

where l_F and l_G are slowly varying. It is then easy to prove that

$$\Lambda_F^-(x) = x^{\theta_X}\bar{l}_F(x),\ \Lambda_G^-(x) = x^{\theta_C}\bar{l}_G(x),\ \Lambda_H(x) = x^{1/\theta_Z} l_H(x), \Lambda_H^-(x) = x^{\theta_Z} l(x) \text{ and } \Lambda_F \circ \Lambda_H^-(x) = x^a \tilde{l}(x),$$

where $\theta_Z = \min(\theta_X, \theta_C)$, $a = \theta_Z/\theta_X$, and $\bar{l}_F$, $\bar{l}_G$, l and $\tilde{l}$ are slowly varying. More precisely, we have the following Lemma, under the second order condition (10).

Lemma 2. *Under* (2), (3) *and* (10), *we have,*

$$\begin{array}{rlrcl} l_F(x) = c_F(1 - x^{\rho_F} v(x)) & & \text{and} & l_G(x) & = c_G(1 - x^{\rho_G} v(x)), \\ \bar{l}_F(x) = c_F^{-\theta_X}(1 - x^{\theta_X \rho_F} v(x)) & & \text{and} & \bar{l}_G(x) & = c_G^{-\theta_C}(1 - x^{\theta_C \rho_G} v(x)), \\ l_H(x) = c_H(1 - x^{\rho_H} v(x)),\ l(x) = c_H^{-\theta_Z}(1 - x^{\rho} v(x)) & & \text{and} & \tilde{l}(x) & = \tilde{c}(1 - x^{\tilde{\rho}} v(x)), \end{array}$$

for different slowly varying functions generically noted v, *with*

$$c_H = \begin{cases} c_F & \text{if } \theta_X < \theta_C \\ c_F + c_G & \text{if } \theta_X = \theta_C \\ c_G & \text{if } \theta_X > \theta_C \end{cases}, \quad \tilde{c} = c_H^{-a} c_F = \begin{cases} 1 & \text{if } \theta_X < \theta_C \\ c_F/(c_F + c_G) & \text{if } \theta_X = \theta_C \\ c_G^{-a} c_F & \text{if } \theta_X > \theta_C \end{cases},$$

$$\rho_H = \begin{cases} \max(\rho_F, 1/\theta_C - 1/\theta_X) & \text{if } \theta_X < \theta_C \\ \max(\rho_F, \rho_G) & \text{if } \theta_X = \theta_C \\ \max(\rho_G, 1/\theta_X - 1/\theta_C) & \text{if } \theta_X > \theta_C \end{cases}, \quad \rho = \theta_Z \rho_H = \begin{cases} \max(\theta_X \rho_F, d - 1) & \text{if } \theta_X < \theta_C \\ \max(\theta_X \rho_F, \theta_X \rho_G) & \text{if } \theta_X = \theta_C \\ \max(\theta_C \rho_G, a - 1) & \text{if } \theta_X > \theta_C \end{cases},$$

and

$$\tilde{\rho} = \begin{cases} \rho & \text{if } \theta_X \leqslant \theta_C \\ \max(\theta_C \rho_G, \theta_C \rho_F, a - 1) & \text{if } \theta_X > \theta_C \end{cases}.$$

The proof of this Lemma is based on Theorem B.2.2 in de Haan and Ferreira (2006) as well as the concept of de Bruyn conjugate (see Proposition 2.5 in Beirlant et al. (2004)). Details are ommited for brevity.

Remark 5. *It is clear that all the aforementioned slowly varying functions satisfy the second order condition SR2 with the corresponding second order parameters defined in the previous Lemma. In particular, rate functions* B *and* $\tilde{B}$ *associated, respectively, to* l *and* $\tilde{l}$ *satisfy* $\frac{x^{\tilde{\rho}} v(x)}{\tilde{B}(x)} \to -1/\tilde{\rho}$ *and* $\frac{x^{\rho} v(x)}{B(x)} \to -1/\rho$, *as* $x \to +\infty$, *with* v, *the appropriate slowly varying function (see again Theorem B.2.2 in de Haan and Ferreira (2006)) .*

Remind that the function p has been defined by $p(x) = \mathbb{P}(\delta = 1 | Z = x)$ at the start of the proof of Lemma 1. The following Lemma provides useful developments of p and $p \circ \Lambda_H^-$. In particular, it provides details about the rate of convergence of $p(x)$, as $x \to +\infty$. Its proof is based on the fact that

$$p(x) = \frac{\bar{G}(x) f(x)}{\bar{G}(x) f(x) + \bar{F}(x) G(x)},$$

where f and g are respectively the derivatives of F and G, as well as on the results of Lemma 2. It is omitted for brevity.

Lemma 3. *Under* (2), (3) *and* (10), *we have*

$$\frac{1}{p(x)} = 1 + \frac{\theta_X}{\theta_C} x^{\frac{1}{\theta_C} - \frac{1}{\theta_X}} \frac{l_G(x)}{l_F(x)} (1 + o(1)).$$

In particular, as $x \to +\infty$,

$$p(x) \to \begin{cases} 1 & \text{if } \theta_X < \theta_C, \\ \tilde{c} = c_F/(c_F + c_G) & \text{if } \theta_X = \theta_C, \\ 0 & \text{if } \theta_X > \theta_C. \end{cases}$$

Moreover, we have

$$\begin{aligned}
\text{if } \theta_X < \theta_C, \quad 1/(p \circ \Lambda_H^-)(x) &= 1 + d\frac{c_G}{c_F^d} x^{d-1}(1 - x^{-\beta} v(x)) \quad , \\
\text{if } \theta_X = \theta_C, \quad (p \circ \Lambda_H^-)(x) &= \tilde{c}(1 - x^{\rho} v(x)), \\
\text{if } \theta_X > \theta_C, \quad 1/(p \circ \Lambda_H^-)(x) &= 1 + \frac{1}{a\tilde{c}} x^{1-a}(1 - x^{\tilde{\rho}} v(x)),
\end{aligned}$$

where $d = \theta_X/\theta_C$, v *is a generic notation for a slowly varying function and*

$$-\beta = \max(\theta_X \rho_F, \theta_X \rho_G, d - 1).$$

6.2. Technical Lemmas

Let $E_1, \ldots, E_n$ be n iid standard exponential random variables.

Lemma 4. *According to Lemma 1.4.3. in Reiss (1989), we have*

$$(E_{n-j+1,n} - E_{n-k,n})_{1 \leqslant j \leqslant k} \stackrel{d}{=} (\tilde{E}_{k-j+1,k})_{1 \leqslant j \leqslant k},$$

where $\tilde{E}_1, \ldots, \tilde{E}_k$ *are* k *independent standard exponential random variables.*

Lemma 5. *Under condition* H_1*, we have, as* $n \to +\infty$*,*

$$\frac{E_{n-k,n}}{L_{nk}} \xrightarrow{\mathbb{P}} 1, \ \frac{E_{n-j+1,n}}{\log(n/i)} \xrightarrow{\mathbb{P}} 1, \ \textit{uniformly on } j = 1, \ldots k \ \textit{ and } \ \sqrt{k}(E_{n-k,n} - L_{nk}) \xrightarrow{d} N(0,1).$$

We refer to Girard (2004,b) for the proof of this Lemma.

Lemma 6. *If we consider the classical random censoring model* (1) *with continuous distribution functions* F *and* G *of the variables* X *and* C*, then the following in-probability results hold :*

$$\begin{aligned}
&\left|\hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \Lambda_F(Z_{n-j+1,n})\right| = O_{\mathbb{P}}(1/\sqrt{j-1}), \ \textit{for } j = 2, \ldots, k+1, \\
&\left|\hat{\Lambda}_{nF}(Z_{n,n}) - \Lambda_F(Z_{n,n})\right| = O_{\mathbb{P}}(1).
\end{aligned}$$

The first statement is a part of Theorem 1 in Csorgo (1996). For the second statement, one has to make a careful examination of Theorem 2.1 in Zhou (1991), in a narrower context, since the samples (X_i) and (C_i) we consider are i.i.d. , whereas Zhou considers possibly non-identically distributed censoring variables C_i. In pages 2269-2270 of the mentioned paper, one can find out that the maximum observed value (named T_n) does not have to be excluded from the probability bound (2.3) : it can indeed be proved, by following the steps of the proof of (2.3), that for every n,

$$\forall \epsilon > 0, \quad \mathbb{P}\left[\sup_{t \leqslant Z_{n,n}} \left|\hat{\Lambda}_{nF}(t) - \Lambda_F(t)\right| > \epsilon\right] \leqslant 6\epsilon^{2/3}.$$

So the second statement of Lemma 6 follows.

6.3. Treatment of the remainder or bias terms $R_{i,k}^{(\Delta)}$*, related to the main statistic* Δ_n

These terms appear in the representation (16) of Δ_n. We have the following results :

Lemma 7. *Under the assumptions of Theorem 1, as* n *tends to infinity,*

$$\sqrt{k} L_{nk}^{1-b} R_{i,k}^{(\Delta)} \xrightarrow{\mathbb{P}} 0, \ \textit{for } j \in \{1, 2, 5, 6\}$$

$$\sqrt{k} L_{nk}^{1-b} R_{3,k}^{(\Delta)} \xrightarrow{\mathbb{P}} -\frac{\tilde{\alpha}}{\rho} \ \textit{if } \theta_X < \theta_C \ \textit{and } 0 \ \textit{if } \theta_X \geqslant \theta_C.$$

$$\sqrt{k} L_{nk}^{1-b} R_{4,k}^{(\Delta)} \xrightarrow{\mathbb{P}} -\tilde{\alpha} \ \textit{if } \theta_X < \theta_C \ \textit{and } 0 \ \textit{if } \theta_X \geqslant \theta_C.$$

Proof of Lemma 7

- Remind that

$$\begin{aligned} R_{1,k}^{(\Delta)} &= \Delta_n - \frac{1}{k}\sum_{j=1}^{k}\left(\frac{\hat{\Lambda}_{nF}(Z_{n-j+1,n})}{\Lambda_F(Z_{n-j+1,n})}\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})} - 1\right) \\ &= \frac{1}{k}\sum_{j=1}^{k}\left(\log(1+\xi_{j,n}) - \xi_{j,n}\right), \end{aligned}$$

where

$$\xi_{j,n} = \frac{\hat{\Lambda}_{nF}(Z_{n-j+1,n})}{\Lambda_F(Z_{n-j+1,n})}\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})} - 1$$

Introducing $\Delta_j = \hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \Lambda_F(Z_{n-j+1,n})$, for $j = 1, \ldots, k+1$, we have readily

$$\xi_{j,n} = \frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}\left(\Delta_j\frac{\Lambda_F(Z_{n-k,n})}{\Lambda_F(Z_{n-j+1,n})} - \Delta_{k+1}\right)\frac{1}{\Lambda_F(Z_{n-k,n})}.$$

But Lemma 6 implies that $|\Delta_j| = O_{\mathbb{P}}(1/\sqrt{j-1})$ for all $j = 2, \ldots, k+1$, $|\Delta_1| = O_{\mathbb{P}}(1)$ and $\frac{\Lambda_F(Z_{n-k,n})}{\hat{\Lambda}_{nF}(Z_{n-k,n})}$ tends to 1, in probability. Let $E_1, \ldots, E_n$ be n independent standard exponential random variable such that $\frac{1}{\Lambda_F(Z_{n-k,n})} = \frac{E_{n-k,n}^{-a}}{\tilde{l}(E_{n-k,n})}$, where $\tilde{l}$ tends to $\tilde{c}$ at infinity. Moreover, $\frac{\Lambda_F(Z_{n-k,n})}{\Lambda_F(Z_{n-j+1,n})} \leqslant 1$ and $\frac{E_{n-k,n}}{L_{nk}}$ tends to 1 (see Lemma 5). Thus, we obtain $|\xi_{1,n}| \leqslant (1+o_{\mathbb{P}}(1))\left(O_{\mathbb{P}}(1) + O_{\mathbb{P}}(1/\sqrt{k})\right)L_{nk}^{-a}(1/\tilde{c} + o_{\mathbb{P}}(1))$ and

$$|\xi_{j,n}| \leqslant (1+o_{\mathbb{P}}(1))\left(O_{\mathbb{P}}(1/\sqrt{j-1}) + O_{\mathbb{P}}(1/\sqrt{k})\right)L_{nk}^{-a}(1/\tilde{c} + o_{\mathbb{P}}(1)), \text{ for } j = 2, \ldots, k.$$

Therefore $\xi_{1,n}^2 \leqslant O_{\mathbb{P}}(1)L_{nk}^{-2a}$ and

$$\xi_{j,n}^2 \leqslant O_{\mathbb{P}}(1)\frac{L_{nk}^{-2a}}{j-1} \text{ for } j = 2\ldots, k.$$

Consequently, since $a > 0$, $\sup_{1\leqslant j\leqslant k}|\xi_{j,n}|$ tends to 0, in probability, and thus, using the inequality $0 \leqslant x - \log(1+x) \leqslant x^2$ $(\forall x \geqslant -1/2)$, we obtain,

$$0 \leqslant -R_{1,k}^{(\Delta)} \leqslant \frac{1}{k}\sum_{j=1}^{k}\xi_{j,n}^2.$$

But $\frac{1}{k}\sum_{j=1}^{k}1/j \sim \frac{\log k}{k}$. Hence

$$0 \leqslant -\sqrt{k}L_{nk}^{1-b}R_{1,k}^{(\Delta)} \leqslant O_{\mathbb{P}}(1)\frac{\log k}{\sqrt{k}}L_{nk}^{1-b-2a}.$$

Let $\epsilon > 0$. We have $1-b-2a = 3b-1$, and so we want

$$\sqrt{k}(\log k)^{-1}L_{nk}^{1-3b} = (k^{\epsilon}/\log k)\left(\sqrt{k}L_{nk}^{(1-3b)/(1-2\epsilon)}\right)^{1-2\epsilon}$$

to go to $+\infty$. This is automatic when $0 \leqslant b \leqslant 1/3$. If $b > 1/3$ (*i.e.* when $\theta_X > 3\theta_C$), we can write $(1-3b)/(1-2\epsilon) = 1-3b-\delta$ for some positive δ and small enough ϵ, and we have $\sqrt{k}L_{nk}^{1-3b+\delta} = \sqrt{k}L_{nk}^{-b} \times L_{nk}^{-2b+1-\delta}$: the first factor goes to infinity (it is the CLT rate, assumption $H_4(i)$), and the second factor as well for δ (*i.e.* ϵ) small enough because b is always smaller than 1/2.

- Remind that

$$R_{2,k}^{(\Delta)} = \frac{1}{\Lambda_F(Z_{n-k,n})}\frac{1}{k}\sum_{j=1}^{k}\left(\hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \Lambda_F(Z_{n-j+1,n})\right)\left(\frac{\Lambda_F(Z_{n-k,n})}{\Lambda_F(Z_{n-j+1,n})} - 1\right)$$

and that $\frac{\Lambda_F(Z_{n-k,n})}{\Lambda_F(Z_{n-j+1,n})} = x_{j,n}^{-a}\frac{\tilde{l}(E_{n-k,n})}{\tilde{l}(E_{n-j+1,n})}$, where $x_{j,n} = \frac{E_{n-k,n}}{E_{n-j+1,n}} \to 1$, uniformly on j (see Lemma 5). Hence, using the fact that $\sup_{1\leqslant j\leqslant k}|\hat{\Lambda}_{nF}(Z_{n-j+1,n}) - \Lambda_F(Z_{n-j+1,n})| = O_{\mathbb{P}}(1)$ (see Lemma 6), we obtain

$$|R_{2,k}^{(\Delta)}| \leqslant O_{\mathbb{P}}(1)\frac{E_{n-k,n}^{-a}}{\tilde{l}(E_{n-k,n})}\left(\frac{1}{k}\sum_{j=1}^{k}|x_{j,n}^{-a} - 1| + \frac{1}{k}\sum_{j=1}^{k}x_{j,n}^{-a}\left|\frac{\tilde{l}(E_{n-k,n})}{\tilde{l}(E_{n-j+1,n})} - 1\right|\right).$$

Introducing, once again, $\tilde{E}_1, \ldots, \tilde{E}_k$, k independent standard exponential random variables, such that, $\frac{E_{n-j+1,n}-E_{n-k,n}}{E_{n-k,n}} \stackrel{d}{=} \frac{\tilde{E}_{k-j,k}}{E_{n-k,n}}$ (see Lemma 4), and using a Taylor expansion, we have

$$|R_{2,k}^{(\Delta)}| \leqslant O_{\mathbb{P}}(1) E_{n-k,n}^{-a} \left(\frac{1}{k} \sum_{j=1}^{k} \frac{\tilde{E}_{k-j,k}}{E_{n-k,n}} + \frac{1}{k} \sum_{j=1}^{k} \left| \frac{\tilde{l}(E_{n-k,n})}{\tilde{l}(E_{n-j+1,n})} - 1 \right| \right).$$

Since $\bar{E}_n = \frac{1}{k}\sum_{j=1}^{k} \tilde{E}_j$ and $\frac{E_{n-k,n}}{L_{nk}}$ tend to 1, in probability, the first term of the right hand side multiplied by $\sqrt{k} L_{nk}^{1-b}$ tends to 0, by the fact that $\sqrt{k} L_{nk}^{-a-b}$ tends to 0 under condition $H_2(iii)$, $H_3(ii)$ or $H_4(iv)$. For the second term of the right hand side, we proceed as for $R_{n,\tilde{l}}$ (see the proof of Proposition 2), by using the fact that condition $R_{\tilde{l}}(\tilde{B}, \tilde{\rho})$ implies $R_{1/\tilde{l}}(-\tilde{B}, \tilde{\rho})$ and again that $\sqrt{k} L_{nk}^{-a-b}$ tends to 0.

- Remind that

$$R_{3,k}^{(\Delta)} = \frac{\hat{p}_k}{E_{n-k,n}^{a}} \left(\frac{1}{\tilde{l}(E_{n-k,n})} - \frac{1}{\tilde{c}} \right),$$

where, according to Lemma 2, we have $1 - \frac{\tilde{l}(x)}{\tilde{c}} = x^{\tilde{\rho}} v(x)$, with v slowly varying. Hence,

$$R_{3,k}^{(\Delta)} = (1 + o_{\mathbb{P}}(1)) E_{n-k,n}^{-a} \frac{\hat{p}_k}{\tilde{c}} E_{n-k,n}^{\tilde{\rho}} v(E_{n-k,n}).$$

We prove, in Lemma 1, that $L_{nk}^{1-a} \frac{\hat{p}_k}{\tilde{c}}$ tends to a. Moreover, since v is slowly varying and $\frac{E_{n-k,n}}{L_{nk}}$ tends to 1 (see Lemma 5), we obtain

$$\sqrt{k} L_{nk}^{1-b} R_{3,k}^{(\Delta)} = a(1 + o_{\mathbb{P}}(1)) \sqrt{k} L_{nk}^{-b+\tilde{\rho}} v(L_{nk}).$$

This term tends to 0 in the case $\theta_X \geqslant \theta_C$, under condition $H_3(i)$ or $H_4(ii)$. In the case $\theta_X < \theta_C$, we use the fact that $\frac{x^{\tilde{\rho}} v(x)}{\tilde{B}(x)} \to -\frac{1}{\tilde{\rho}}$ (see Remark 5 in the Appendix). Thus,

$$\sqrt{k} L_{nk}^{1-b} R_{3,k}^{(\Delta)} = -\frac{1}{\tilde{\rho}} (1 + o_{\mathbb{P}}(1)) \sqrt{k} L_{nk}^{-b} \tilde{B}(L_{nk}),$$

which tends to $-\frac{\tilde{\alpha}}{\rho}$ under condition $H_2(ii)$, since $\rho = \tilde{\rho}$, in this case.

- Remind that

$$R_{4,k}^{(\Delta)} = -\frac{1}{k} \sum_{j=1}^{k} \left(\frac{E_{n-j+1,n}}{E_{n-k,n}} \right)^{a} \left(\frac{\tilde{l}(E_{n-j+1,n})}{\tilde{l}(E_{n-k,n})} - 1 \right).$$

The treatment of this term is very similar to that of $R_{n,\tilde{l}}$ (see the proof of Proposition 2). It relies on condition $R_{\tilde{l}}(\tilde{B}, \tilde{\rho})$, as well as $H_2(ii)$, $H_3(i)$ or $H_4(ii)$. It is thus omitted.

- Remind that

$$R_{5,k}^{(\Delta)} = -\frac{1}{k} \sum_{j=1}^{k} \left\{ \left(\left(1 + \frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}} \right)^{a} - 1 \right) - a \frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}} \right\}.$$

This term is 0 in the case $\theta_X \leqslant \theta_C$ ($a = 1$). So, we only consider the case $\theta_X > \theta_C$ (where $0 < a < 1$). It is clear (see Lemmas 4 and 5) that $\xi_{j,n} = \frac{\tilde{E}_{k-j+1,k}}{E_{n-k,n}} \stackrel{d}{=} \frac{E_{n-j+1,n}}{E_{n-k,n}} - 1$ tends to 0, uniformly in j. Hence, by a Taylor expansion, we obtain

$$\begin{aligned} R_{5,k}^{(\Delta)} &= -(1 + o_{\mathbb{P}}(1)) \tfrac{1}{k} \textstyle\sum_{j=1}^{k} \frac{a(a-1)}{2} \xi_{j,n}^2 \\ &\stackrel{d}{=} (1 + o_{\mathbb{P}}(1)) \tfrac{a(1-a)}{2} \tfrac{1}{E_{n-k,n}^2} \tfrac{1}{k} \textstyle\sum_{j=1}^{k} \tilde{E}_j^{\,2} \sim \tfrac{a(1-a)}{2} L_{nk}^{-2}, \text{ (in probability)}, \end{aligned}$$

and we conclude using $H_4(iv)$.

- Finally, remind that

$$R_{6,k}^{(\Delta)} = \frac{\hat{p}_k}{\tilde{c} E_{n-k,n}} \left(E_{n-k,n}^{1-a} - L_{nk}^{1-a} \right).$$

This term is 0 in the case $\theta_X \leqslant \theta_C$ ($a = 1$). So, we only consider the case $\theta_X > \theta_C$, where $0 < a < 1$ and $\hat{p}_k$ tends to 0 (see Lemma 1 in Subsection 3.1). By the mean value theorem,

$$E_{n-k,n}^{1-a} - L_{nk}^{1-a} = (1-a) L_{nk}^{-a} \left(\frac{\widetilde{L}_{nk}}{L_{nk}} \right)^{-a} (E_{n-k,n} - L_{nk}),$$

where $\widetilde{L}_{nk}$ is between L_{nk} and $E_{n-k,n}$. Hence $\frac{\widetilde{L}_{nk}}{L_{nk}}$ tends to 1 and, since $\sqrt{k}(E_{n-k,n} - L_{nk}) \xrightarrow{d} N(0,1)$ (see Lemma 5), we have

$$\sqrt{k} L_{nk}^{1-b} |R_{6,k}^{(\Delta)}| \leqslant o_{\mathbb{P}}(1) L_{nk}^{-b-a} = o_{\mathbb{P}}(1).$$