Extreme value analysis with non-stationary observations*

Laurens de Haan

University of Lisbon and Erasmus University Rotterdam

Chen Zhou

De Nederlandsche Bank and Erasmus University Rotterdam

August 27, 2012

Abstract We consider extreme value analysis based on independent observations from non-identical distributions with comparable tails. We first establish the asymptotic behavior on the weighted tail empirical process based on the non-stationary observations. As an application, we show that if the tail distributions of the observations are heavy-tailed, then the asymptotic normality of the Hill estimators remain valid even if the observations are non-stationary.

Keywords: Tail empirical process, weighted approximation, the Hill estimator.

AMS Classification Numbers: ...

*First draft

1 Introduction

Extreme value theory, the probability theory of rare events, is usually developed for independently and identically distributed (i.i.d.) random variables. Correspondingly, extreme value analysis, i.e. statistical inference on the tail region of a distribution function, is usually established for i.i.d. observations. Leadbetter et al. (1983), Hsing (1991) and Drees (2000) developed extreme value theory for stationary weakly dependent time series. The latter two references also established statistical tools for that situation. For independent but not identically distributed random variables, a basic probabilistic result is in Mejzler (1956). In the present paper we establish statistical tools for a class of independent, not identically distributed observations.

As in the i.i.d. case, a basic tool for establishing statistical theory is a weighted approximation of the tail empirical process by Brownian motion (c.f. Einmahl (1997) for the i.i.d. case). In the case of non-identically distributed observations, this approximation is known for the entire empirical process (Shorack and Wellner (1986), Chapter 25), but not for the tail empirical process. The first (the entire empirical process) does not imply the latter (the tail empirical process); see Remark 2.2 below. We shall develop the tail empirical process result that is missing.

The special feature of the non-stationarity we deal with is explained as follows. Suppose independent observations $X_1, \cdots, X_n$ are generated from continuous distribution functions $F_{n,1}, \cdots, F_{n,n}$ with a common right endpoint $x^* = \sup\{x : F_{n,i}(x) < 1\}$ for all $i = 1, 2, \cdots, n$. Assume that there exists a distribution function F, with the same right endpoint x^*, such that

$$\lim_{x \to x^*} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c_{n,i} \tag{1.1}$$

holds uniformly for all n and all $1 \leq i \leq n$. Here x^* can be either a finite real number or $+\infty$. Moreover, we assume that the constants $\{c_{n,i}\}_{i=1}^n$ are bounded away from zero and infinity, i.e. there exist a and b such that $0 < a \leq c_{n,i} \leq b < +\infty$ for all n and all $1 \leq i \leq n$. Notice that the distribution function F and the constants $\{c_{n,i}\}_{i=1}^n$ are not uniquely identified unless a normalization condition is imposed. We impose the following normalization condition:

$$\frac{1}{n}\sum_{i=1}^{n} c_{n,i} = 1. \tag{1.2}$$

With this normalization condition, it is clear that as $x \to x^*$,

$$1 - F(x) \sim \frac{1}{n}\sum_{i=1}^{n}(1 - F_{n,i}(x)).$$

Hence, in the tail region, the distribution function F can be regarded as an "average" distribution function.

The condition (1.1) is a simple model for non-stationary observations because it assumes comparability of the tail parts of the underlying distribution functions only, while no assumption on the middle parts of the distributions has been made. We consider statistical inference in this framework. Also we consider large sample sizes n, i.e. $n \to \infty$. We further assume that the common distribution function F does not depend on the sample size n. Such an assumption is crucial for establishing asymptotic theory. Nevertheless, since potential applications are based on a specific finite n, it can be always regarded as a valid assumption.

The tail empirical process result will be proved first for a simple case with uniform distribution function (Theorem 2.1). Then, generalizing the result to distributions with comparable tails as given in (1.1) and (1.2), we obtain a general asymptotic theory on the weighted tail empirical process based on those non-stationary observations (Theorem 2.3).

Next, we use the tail empirical process tool to study extreme value analysis based on non-stationary observations. For that purpose, we consider observations drawn from distributions in the domain of attraction of an extreme value distribution. With the notation $U(x) := \left(\frac{1}{1-F}\right)^{\leftarrow}(x)$, where $^{\leftarrow}$ denotes the left-continuous inverse function, the domain of attraction condition on the F can be written as follows: there exists a real number γ, the extreme value index, and a positive function $a(t)$ such that, for all $x > 0$,

$$\lim_{t \to +\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \tag{1.3}$$

The case $\gamma > 0$ corresponds to the so-called "heavy-tailed" case, while in the case $\gamma < 0$, $U(+\infty) < \infty$ is the finite right endpoint x^* of the distribution function F. By combining the condition (1.1) with the domain of attraction condition, we show that all distribution functions $F_{n,i}$ have the same extreme value index. In other words, the extreme value index of the "average" distribution function also indicates the shape of the tail distribution for each individual observation. Thus, estimating γ is important in making statistical inference on the tail properties of the data.

In the case that the observations $X_1, \cdots, X_n$ are i.i.d. with a heavy-tailed distribution function F, i.e. $\gamma > 0$, the extreme value index γ can be estimated by the Hill estimator in Hill (1975) as follows.

$$\hat{\gamma}_H(s_n) := \frac{\sum_{i=1}^n (\log X_i - \log s_n)_+}{\sum_{i=1}^n 1_{X_i > s_n}}, \tag{1.4}$$

where s_n is a threshold such that $s_n \to \infty$ and $n(1 - F(s_n)) \to \infty$ as $n \to \infty$. Instead of taking an deterministic threshold s_n, an alternative and usual procedure is to take a stochastic threshold which is a high order statistic. Rank the observations $X_1, \cdots, X_n$ as $X_{n,n} \geq X_{n,n-1} \geq \cdots \geq X_{n,1}$. Then, the $(n-k)$-th order statistic $X_{n,n-k}$ can be a suitable threshold provided that as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. Subsequently, the Hill estimator turns to be

$$\hat{\gamma}_{H,k} := \frac{1}{k} \sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k}. \tag{1.5}$$

In the i.i.d. case, the consistency of the two Hill estimators, (1.4) and (1.5) is proved under the domain of attraction condition (1.3) with $\gamma > 0$. Under a proper second order condition quantifying the speed of convergence in condition (1.3), the asymptotic normality for the two Hill estimators has been established for i.i.d. observations in Goldie and Smith (1987) and Davis and Resnick (1984) respectively. Note that the proof of the asymptotic normality in that case can also be derived from the asymptotic behavior of the tail empirical process, see Example 5.1.5 in de Haan and Ferreira (2006).

In the case that the observations are not i.i.d., we use the same definition of the Hill estimator with a deterministic threshold as in (1.4). For defining the Hill estimator with a stochastic threshold as in (1.5), the threshold is again chosen as the $k+1$-th highest value among $X_1, \cdots, X_n$, $X_{n,n-k}$. Notice that this is different from the usual definition of order statistic: the observations are now drawn from different distributions. Nevertheless, we show that $X_{n,n-k}$ is a good estimator of $U(n/k)$, the $(1-k/n)$ quantile of the "average" distribution function $F(x)$. As our main result, we prove that the asymptotic normality results of the Hill estimators with deterministic threshold and stochastic threshold remain valid. As a side result, we prove that by taking $s_n = U(n/k)$, we have that

$$\sqrt{k}\left(\hat{\gamma}_H\left(U\left(\frac{n}{k}\right)\right) - \hat{\gamma}_{H,k}\right) \xrightarrow{P} 0,$$

as $n \to \infty$. Such a result seems to be new even in the i.i.d. context.

The paper is organized as follows. Section 2 studies the tail empirical process based on non-stationary observations. Section 3 proves the asymptotic normality of the Hill estimators based on non-stationary observations for the $\gamma > 0$ case. Section 4 concludes and discusses further extensions. The appendix provides proofs of an auxiliary theorem.

2 Tail empirical process with non-stationary observations

In this section, we investigate the asymptotic behavior of the tail empirical process based on non-stationary observations. Notice that throughout this section, we do not impose any assumption on the tail of the average distribution function F. In other words, results in this section do not require that F belongs to the domain of attraction.

The tail empirical process is defined as follows. Firstly, denote the "empirical distribution function" based on the non-stationary observations as follows:

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \le x}, \;\; x < x^*. \tag{2.1}$$

Consider the tail region $x \ge s_n$, where $\{s_n\}_{n=1}^{\infty}$ is a series of real numbers such that

$s_n \to x^*$ and $n(1-F(s_n)) \to \infty$ as $n \to \infty$. Notice that, with denoting $k = n(1-F(s_n))$, the requirement on the s_n series is equivalent to $s_n = U\left(\frac{n}{k}\right)$ for some $k := k(n)$ such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. The tail empirical process is defined as

$$\mathbb{F}_n(x) := \sqrt{\frac{n}{1-F(s_n)}}\left((1-\hat{F}(x)) - (1-F(x))\right) = \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{X_i > x} - \frac{n}{k}(1-F(x))\right),$$

for $x > s_n$ or $x > U(n/k)$.[1]

To establish the asymptotic theory on the process $\mathbb{F}_n(x)$ for $x > s_n$, we start with a simple case by assuming that F is the standard uniform distribution and that the asymptotic relation in (1.1) appears as an exact relation, i.e. $1 - F_{n,i}(x) = c_{n,i}(1-x)$ for $1 - 1/c_{n,i} \leq x \leq 1$. In other words, X_i is a uniformly distributed random variable on $[1-1/c_{n,i}, 1]$. Hence $\eta_i = c_{n,i}(1-X_i)$ follows a standard uniform distribution. From the independence among the X_i, the random variables $\eta_1, \cdots, \eta_n$ are i.i.d.. Hence, with the notation η_i and $t = (1-x)\frac{n}{k}$, the process $\mathbb{F}_n$ in this simple case can be rewritten as

$$\mathbb{S}_n(t) := \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{\eta_i < c_{n,i} t k/n} - t\right),$$

with $t \in [0,1]$.

We remark that the process $\mathbb{S}_n(t)$ is comparable with the usual tail empirical process based on i.i.d standard uniform random variables η_i as follows. In the classical empirical process theory, the process $\mathbb{G}_n(t) := \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} 1_{\eta_i < t} - t\right)$ weakly converges to a standard Brownian bridge $B(t)$ in the $D-$space equipped with norm $\| \cdot / q \|$, where q is a proper weight function, see, e.g. Shorack and Wellner (1986). The tail empirical process is obtained from the process $\mathbb{G}_n(t)$ by letting t shrink to zero at a proper speed as $n \to \infty$. More specifically, suppose that a series of positive numbers $k := k(n)$ satisfies that $k \to +\infty$ and $k/n \to 0$ as $n \to \infty$. The tail empirical process is defined as

$$\tilde{\mathbb{G}}_n(t) := \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{\eta_i < tk/n} - t\right),$$

[1] Empirical processes are usually defined by scaling up the difference $\hat{F}(x) - F(x)$. Here we choose to compare $1 - \hat{F}(x)$ with $1 - F(x)$ because we focus on the right tail of the distribution.

for $0 < t \leq 1$. In the i.i.d. case, the tail empirical process $\tilde{\mathbb{G}}_n(t)$ converges weakly to a standard Brownian motion $W(t)$ in the $D-$space equipped with norm $\| \cdot/q \|$, see, e.g. Einmahl (1997). Here the conditions on the weight function q may differ from those used for the empirical process. More specifically, the weight function q for the tail empirical process can be chosen from the set $\mathbb{Q}$ defined by

$$\mathbb{Q} = \left\{ q : [0,2] \to [0,+\infty) : \ q \text{ continuous and increasing}, \ \frac{q(u)}{\sqrt{u}} \text{decreasing}, \ \int_0^2 q(u)^{-2} du < \infty \right\}.$$

An example of such a q function is $q(t) = t^{1/2-\varepsilon}$, with $0 < \varepsilon < 1/2$. Throughout the paper, we do use this q function, only except for the general result in Theorem 2.1.

It is clear that, when considering the i.i.d. case the process $\mathbb{S}_n(t)$ coincides with the tail empirical process $\tilde{\mathbb{G}}_n(t)$. For the non-identically distributed case, it differs from $\tilde{\mathbb{G}}_n(t)$ only by allowing different thresholds for different η_i.

The following theorem shows that, similar to $\tilde{\mathbb{G}}_n(t)$, under the condition (1.2), $\mathbb{S}_n(t)$ weakly converges to a standard Brownian motion in the $D-$space equipped with a norm $\| \cdot/q \|$, for any $q \in \mathbb{Q}$. The proof is postponed to the Appendix.

Theorem 2.1 *Suppose a series of positive numbers $k := k(n)$ satisfies that $k \to +\infty$ and $k/n \to 0$ as $n \to \infty$. Suppose the weight function q belongs to the class $\mathbb{Q}$. For each n, there exists a standard Brownian motion $W_n(t)$ defined on the same probability space as $(\eta_1, \cdots, \eta_n)$ such that as $n \to \infty$,*

$$\sup_{0 \leq t \leq 2} \frac{1}{q(t)} \left| \mathbb{S}_n(t) - W_n(t) \right| \xrightarrow{P} 0$$

Remark 2.2 *In Chapter 25 of Shorack and Wellner (1986), the asymptotic theory on empirical process with non-identically distributed observations has been established. The theory shows weak convergence in the $D-$space equipped with a norm $\| \cdot/q \|$ for some weight function q. At a first glance, it seems that the result in Theorem 2.1 can be derived from there. This is in fact not possible. The reason is similar to the situation in the i.i.d. case: the asymptotic property on the empirical process with a weight function is not sufficient for deriving the asymptotic property on the tail empirical process with a* *weight function. The details are as follows.*

We have that the empirical process $\mathbb{G}_n(t)$ weakly converges to a standard Brownian bridge $B_n(t)$ in the $D-$space equipped with norm $\| \cdot/q \|$, for some weight function q. For example, $q(t) = (t(1-t))^{1/2-\varepsilon}$, for any $\varepsilon < 1/2$. By applying this result with replacing t to tk/n, one can obtain that $\tilde{\mathbb{G}}_n(t)$ weakly converges to a standard Brownian motion $W_n(t)$ in the sense that

$$\sup_{0\le t\le 1}\left(\frac{k}{n}\right)^{\varepsilon}\frac{1}{t^{1/2-\varepsilon}}\left|\tilde{\mathbb{G}}_n(t) - W_n(t)\right| \xrightarrow{P} 0,$$

as $n \to \infty$. This is not sufficient to get the result on the weighted tail empirical process, because the extra "rate" $\left(\frac{k}{n}\right)^{\varepsilon}$ can not be omitted.

On the other hand, a stronger result on the empirical process with both a "weight function" and an extra "rate" (n^{ε}) in Csörgo and Horváth (1993) Theorem 6.2.1 can help to directly establish the result on tail empirical process with both a "weight function" and an extra "rate" (k^{ε}), when the observations are i.i.d., see, e.g. the proof of Lemma 2.4.10 in de Haan and Ferreira (2006). Unfortunately, we do not have a parallel result for the case with non-identically distributed random variables. That is why a detailed proof of Theorem 2.1 is necessary.

Next, we establish the asymptotic behavior of $\mathbb{F}_n$ for a general distribution function F and under the general framework given in (1.1) and (1.2). For that purpose, we need a second order condition quantifying the speed of convergence in (1.1). Suppose there exists a positive function $A(t)$, such that as $t \to \infty$, $A(t)$ is eventually decreasing and $A(t) \to 0$, and as $x \to \infty$,

$$\left|\frac{\frac{1-F_{n,i}(x)}{1-F(x)} - c_{n,i}}{A\left(\frac{1}{1-F(x)}\right)}\right| = O(1) \tag{2.2}$$

holds uniformly for all n and all $1 \le i \le n$. The following theorem gives the asymptotic behavior for the weighted tail empirical process under the second order condition.

Theorem 2.3 *For each n, let $X_1, \cdots, X_n$ be independent random variables following distribution functions $F_{n,1}, \cdots, F_{n,n}$. Suppose there exists a distribution function F such* *that the conditions* (1.1) *and* (1.2) *hold uniformly for all* $1 \le i \le n$*. Moreover, the second order condition* (2.2) *holds with some function* A*. Suppose a series of positive number* $k := k(n)$ *satisfies that* $k \to \infty$*,* $k/n \to 0$ *and* $\sqrt{k}A(n/k) \to 0$ *as* $n \to \infty$*, Then, for each* n *there exists a standard Brownian motion* $W_n(x)$ *defined on* $[0,1]$ *such that, as* $n \to \infty$*, for any* $0 < \varepsilon < 1/2$*,*

$$\sup_{0 \le t \le 1} t^{-1/2+\varepsilon} \left| \mathbb{F}_n \left(U\left(\frac{n}{tk}\right)\right) - W_n(t) \right| \xrightarrow{P} 0. \tag{2.3}$$

To write the theorem in the original form of $\mathbb{F}_n(x)$ on $x \ge s_n$, we use the transformation that $x = U\left(\frac{n}{tk}\right)$, and $s_n = U(n/k)$. It is given as in the following corollary.

Corollary 2.4 *With the same conditions and notations as in Theorem 2.3, for a series of real number* s_n *such that as* $n \to \infty$*,* $s_n \to x^*$*,* $n(1 - F(s_n)) \to +\infty$*, and* $\sqrt{n(1 - F(s_n))}A(1/(1 - F(s_n))) \to 0$*, we have that*

$$\sup_{x \ge s_n} \left(\frac{1 - F(x)}{1 - F(s_n)}\right)^{-1/2+\varepsilon} \left| \mathbb{F}_n(x) - W_n\left(\frac{1 - F(x)}{1 - F(s_n)}\right)\right| \xrightarrow{P} 0.$$

Remark 2.5 *Corollary 2.4 shows that* $\hat{F}(x)$ *is a valid estimator of* $F(x)$ *with proper asymptotic normality. An immediate consequence is that the consistency of such an estimator holds. Here it is meaningless to discuss the consistency in the usual form as* $\hat{F}(x) \xrightarrow{P} F(x)$ *because* $x \ge s_n$ *guarantees that both* $\hat{F}(x)$ *and* $F(x)$ *converges to 1 as* $n \to \infty$*. We consider a stronger form of consistency as* $\frac{1-\hat{F}(x)}{1-F(x)} \xrightarrow{P} 1$ *as* $n \to \infty$*. It can be shown that under the conditions in Theorem 2.3, such a consistency result holds for* $s_n \le x \le t_n$*, with* $t_n = U\left(\sqrt{\frac{n}{1-F(s_n)}}\right)$*. In fact, the conditions can be relaxed in the following direction: the second order condition* (2.2) *and the corresponding condition that* $\sqrt{n(1 - F(s_n))}A(1/(1 - F(s_n))) \to 0$ *as* $n \to \infty$ *can be omitted. The consistency result can be established only under the "first order conditions".*

Nevertheless, the upper bound t_n *is necessary for the consistency result. Because we choose to establish a stronger "consistency" result, such a consistency result can be obtained only if* $1 - F(x)$ *is not too low, or equivalently,* x *is not too close to the end point* x^**. This yields the upper bound of the tail region for which the stronger consistency* *result holds. On the contrary, when studying the asymptotic behavior of the tail empirical process, essentially, we consider the difference between $1-\hat{F}(x)$ and $1-F(x)$ with scaling up by a factor $\sqrt{\frac{n}{1-F(s_n)}}$. Thus an upper bound is not necessary and the relation holds for the entire tail region $x \geq s_n$.*

Proof of Theorem 2.3 Denote $\eta_i = 1-F_{n,i}(X_i)$ for $1 \leq i \leq n$. Then $\eta_1, \cdots, \eta_n$ are i.i.d. random variables from the uniform distribution as those used in defining the process $\mathbb{S}_n$. Then, we have that

$$\begin{aligned}\mathbb{F}_n\left(U\left(\frac{n}{tk}\right)\right) &= \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{X_i>U\left(\frac{n}{tk}\right)} - t\right) \\ &= \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{\eta_i<1-F_{n,i}\left(U\left(\frac{n}{tk}\right)\right)} - t\right)\end{aligned}$$

For the right hand side, with replacing $1-F_{n,i}$ by its approximation, $c_{n,i}(1-F)$, we obtain the process $\mathbb{S}_n$. Hence the general idea of the proof is to approximate $\mathbb{F}_n\left(U\left(\frac{n}{tk}\right)\right)$ by $\mathbb{S}_n(t)$ and to use the asymptotic behavior of $\mathbb{S}_n(t)$ established in Theorem 2.1. To follow such an idea, we need to first evaluate how accurate the approximation is. For that purpose, we establish inequalities stemming from the second order condition to bound the term $\mathbb{F}_n\left(U\left(\frac{n}{tk}\right)\right)$.

The condition (2.2) implies that there exists real numbers $x_0 < x^*$ and $Q > 0$ such that for all $x > x_0$,

$$\left|\frac{\frac{1-F_{n,i}(x)}{1-F(x)} - c_{n,i}}{A\left(\frac{1}{1-F(x)}\right)}\right| < Q$$

holds uniformly for all $1 \leq i \leq n$. Notice that all $c_{n,i}$ are bounded away from zero by $c_{n,i} \geq a > 0$. Hence, for $x > x_0$

$$c_{n,i}\left(1-\frac{Q}{a}A\left(\frac{1}{1-F(x)}\right)\right) < \frac{1-F_{n,i}(x)}{1-F(x)} < c_{n,i}\left(1+\frac{Q}{a}A\left(\frac{1}{1-F(x)}\right)\right).$$

Thus, we get that

$$(1-\delta_n(t))c_{n,i}\frac{tk}{n} < 1-F_{n,i}\left(U\left(\frac{n}{tk}\right)\right) < (1+\delta_n(t))c_{n,i}\frac{tk}{n},$$

where $\delta_n(t) := \frac{Q}{a} A\left(\frac{n}{tk}\right) > 0$. Hence, we get that

$$\begin{aligned}
\mathbb{F}_n\left(U\left(\frac{n}{tk}\right)\right) &= \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{\eta_i < 1 - F_{n,i}\left(U\left(\frac{n}{tk}\right)\right)} - t\right) \\
&\leq \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} 1_{\eta_i < c_{n,i}\frac{t(1+\delta_n(t))k}{n}} - t\right) \\
&= \mathbb{S}_n\left(t(1+\delta_n(t))\right) + \sqrt{k}t\delta_n(t).
\end{aligned}$$

A similar lower bound can be established as

$$\mathbb{F}_n\left(U\left(\frac{n}{tk}\right)\right) \geq \mathbb{S}_n\left(t(1-\delta_n(t))\right) - \sqrt{k}\delta_n(t).$$

Notice that $\delta_n(t) \to 0$ as $n \to \infty$ uniformly for all $t \in [0,1]$. Hence for sufficiently large n, $t(1 \pm \delta_n(t)) \in [0,2]$. Thus, we can apply Theorem 2.1 with $q(t) = t^{1/2-\varepsilon}$ and then replace t with $t(1 \pm \delta_n(t))$. We obtain that, as $n \to \infty$,

$$\sup_{0\leq t\leq 1} t^{-1/2+\varepsilon}\left|\mathbb{S}_n(t(1\pm\delta_n(t))) - W_n(t(1\pm\delta_n(t)))\right| \xrightarrow{P} 0. \tag{2.4}$$

Here we use again the fact that $\delta_n(t) \to 0$ as $n \to \infty$ uniformly for $t \in [0,1]$. Hence, comparing (2.4) with the upper and lower boundaries of the $\mathbb{F}_n\left(U\left(\frac{n}{tk}\right)\right)$ process, for proving the theorem, it suffices to verify the following statements: as $n \to \infty$,

$$\sup_{0\leq t\leq 1} t^{-1/2+\varepsilon}\sqrt{k}t\delta_n(t) \to 0, \tag{2.5}$$

$$\sup_{0\leq t\leq 1} t^{-1/2+\varepsilon}\left|W_n\left(t(1\pm\delta_n(t))\right) - W_n\left(t\right)\right| \xrightarrow{P} 0. \tag{2.6}$$

Notice that $t^{-1/2+\varepsilon}\sqrt{k}t\delta_n(t) = t^{\varepsilon}\frac{Q}{a}\sqrt{kt}A(n/(kt))$. As $n \to \infty$, it is eventually an increasing function with respect to t for $0 \leq t \leq 1$. Hence, $\sup_{0\leq t\leq 1} t^{-1/2+\varepsilon}\sqrt{k}t\delta_n(t) = \sqrt{k}A(n/k) \to 0$ as $n \to \infty$. The relation (2.5) is thus verified.

To prove the relation (2.6), we use the Lévy's modulus of continuity theorem: for a standard Brownian motion $W(t)$, the following relation holds almost surely,

$$\limsup_{\eta\to 0^+} \sup_{0\le s\le t\le 1, t-s\le\eta} \frac{|W(t)-W(s)|}{\sqrt{2\eta\log(1/\eta)}} = 1.$$

Moreover, given $\varepsilon > 0$, for sufficiently small $\eta > 0$, we have that $\sqrt{2\eta\log(1/\eta)} < \eta^{1/2-\varepsilon}$. Therefore, given $\varepsilon > 0$, for sufficiently large n and all $t \le 1$, we have almost surely that

$$|W_n(t(1\pm\delta_n(t))) - W_n(t)| \le (t\delta_n(t))^{1/2-\varepsilon}.$$

which implies that,

$$\sup_{0\le t\le 1} t^{-1/2+\varepsilon} |W_n(t(1\pm\delta_n(t))) - W_n(t)| \le \sup_{0\le t\le 1} t^{-1/2+\varepsilon} \left(t\frac{Q}{a}A\left(\frac{n}{kt}\right)\right)^{1/2-\varepsilon} = \left(\frac{Q}{a}A\left(\frac{n}{k}\right)\right)^{1/2-\varepsilon}.$$

Since $A(n/k) \to 0$ as $n \to \infty$, relation (2.6) is proved. The theorem follows from combining (2.4)-(2.6) with the upper and lower bounds of $\mathbb{F}_n(x)$.□

3 The Hill estimators with non-stationary observations

The results in Section 2 show that when having non-stationary observations drawn from distributions with comparable tails, the tail empirical process has similar asymptotic behavior as that for the i.i.d. case. In this section, we apply this result in extreme value analysis. The main result is Theorem 3.4 on the asymptotic normality of Hill estimators in two forms.

Let $X_1, \cdots, X_n$ be independent random variables following distribution functions $F_{n,1}, \cdots, F_{n,n}$. Suppose there exists a distribution function F such that the conditions (1.1) and (1.2) hold uniformly for all $1 \le i \le n$. To make statistical inference on the tail region of the distribution function F, we further assume that F belongs to the domain of attraction. In other words, with denoting $U(x) = \left(\frac{1}{1-F}\right)^{\leftarrow}(x)$, the condition (1.3) holds for some positive function $a(t)$ and an extreme value index $\gamma \in \mathbb{R}$. The following proposition shows that the distribution function $F_{n,i}$ belongs to the same domain of attraction, for all $1 \le i \le n$.

Proposition 3.1 *Suppose F belongs to the domain of attraction of an extreme value distribution with an extreme value index $\gamma \in \mathbb{R}$. Then each $F_{n,i}$ also belongs to the domain of attraction with the same extreme value index γ.*

Proof of Proposition 3.1 The domain of attraction condition (1.3) is equivalent to the following condition: there exists a positive function $f(t)$ such that, for all x such that $1 + \gamma x > 0$,

$$\lim_{t \to x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}.$$

Together with (1.1), we get that

$$\lim_{t \to x^*} \frac{1 - F_{n,i}(t + xf(t))}{1 - F_{n,i}(t)} = (1 + \gamma x)^{-1/\gamma}.$$

The proposition is thus proved. □

Proposition 3.1 shows that the extreme value index of the "average" distribution function F also characterizes the tail shape of the distribution functions for each observation. Therefore, estimating the extreme value index of F is an important step for making further statistical inference on the underlying dataset.

In the rest of this section, we consider the case that the extreme value index γ is positive. We show that when applying the Hill estimators with non-stationary observations, they are still valid estimates of the extreme value index of the average distribution which possess asymptotic normality.

To establish the asymptotic normality, we need a second order condition on the average distribution function F, see Theorem 3.2.5 in de Haan and Ferreira (2006). Suppose that there exists a function $A(t)$ and a negative real number ρ such that $A(t)$ has either positive or negative sign as $t \to \infty$, $A(t) \to 0$ as $t \to \infty$ and for any $x > 0$,

$$\lim_{t \to \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \qquad (3.1)$$

where $U = (1/(1-F))^{\leftarrow}$ is the quantile function. Here we use the same scaling function $A(t)$ as in (2.2). We first demonstrate that it is a reasonable assumption by the following example.

Consider i.i.d. random variables $Z_1, Z_2 \cdots, Z_n$ following the distribution F which satisfies the second order condition (3.1). Define $X_i = c_{n,i}^{\gamma} Z_i$ for $1 \leq i \leq n$. Then as $x \to \infty$, we have that $F_{n,i}$, the distribution function of X_i, satisfies

$$\lim_{x \to x^*} \frac{1 - F_{n,i}(x)}{1 - F(x)} = \lim_{x \to x^*} \frac{1 - F(x c_{n,i}^{-\gamma})}{1 - F(x)} = c_{n,i}.$$

Hence this is an example of non-stationary observations under the condition (1.1) and (1.2). Moreover, from the second order condition on the U function, we get that

$$\lim_{x \to x^*} \frac{\frac{1-F(tx)}{1-F(x)} - x^{-1/\gamma}}{A\left(\frac{1}{1-F(x)}\right)} = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\gamma\rho},$$

which implies that uniformly

$$\lim_{x \to x^*} \frac{\frac{1-F_{n,i}(x)}{1-F(x)} - c_{n,i}}{A\left(\frac{1}{1-F(x)}\right)} = \lim_{x \to x^*} \frac{\frac{1-F(x c_{n,i}^{-\gamma})}{1-F(x)} - c_{n,i}}{A\left(\frac{1}{1-F(x)}\right)} = c_{n,i} \frac{c_{n,i}^{-\rho} - 1}{\gamma\rho}.$$

Since $c_{n,i}$ are bounded between a and b, we get that the second order condition (2.2) holds with the same scaling function A as that in (3.1). Therefore, it is a reasonable assumption to consider the same A function in the two second order conditions (2.2) and (3.1).

Since F is heavy-tailed, its tail can be approximated by a Pareto distribution. That leads to a specific expression of Theorem 2.3, which is given as in the following proposition.

Proposition 3.2 *Assume that the conditions in Theorem 2.3 hold. Suppose that the second order conditions* (2.2) *and* (3.1) *hold for the same function* $A(t)$*. Then, as* $n \to \infty$,

$$\sup_{u \geq 1} u^{1/(2\gamma)-\varepsilon} \left| \sqrt{k} \left(\frac{n}{k} \left(1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right)\right) - u^{-1/\gamma} \right) - W_n\left(u^{-1/\gamma}\right) \right| \xrightarrow{P} 0. \qquad (3.2)$$

Proof of Proposition 3.2 For any $u \geq 1$, denote $v_n(u) = \frac{n}{k}\left(1 - F\left(uU\left(\frac{n}{k}\right)\right)\right)$. We first show that $v_n(u)$ is approximately $u^{-1/\gamma}$. Notice that the second order condition (3.1) implies that

$$\left| \frac{\frac{1-F(uU(n/k))}{1-F(U(n/k))} \cdot u^{1/\gamma} - 1}{A(n/k)} \right| = O(1)$$

holds uniformly for all $u \geq 1$; see, e.g. inequality (B.3.25) in de Haan and Ferreira (2006). Hence, uniformly for $u \geq 1$,

$$\left| \frac{v_n(u) \cdot u^{1/\gamma} - 1}{A(n/k)} \right| = O(1). \tag{3.3}$$

Next, the relation (2.3) in Theorem 2.3 can be explicitly written as

$$\sup_{0 \leq t \leq 1} t^{-1/2+\varepsilon} \left| \sqrt{k} \left(\frac{n}{k} \left(1 - \hat{F} \left(U \left(\frac{n}{kt} \right) \right) \right) - t \right) - W_n(t) \right| \xrightarrow{P} 0, \tag{3.4}$$

as $n \to \infty$. Since (3.3) implies that $v_n(u)$ is bounded uniformly for all $u \geq 1$, we may and do replace t in (3.4) by $v_n(u)$ and obtain that

$$\sup_{u \geq 1} (v_n(u))^{-1/2+\varepsilon} \left| \sqrt{k} \left(\frac{n}{k} \left(1 - \hat{F} \left(uU \left(\frac{n}{k} \right) \right) \right) - v_n(u) \right) - W_n(v_n(u)) \right| \xrightarrow{P} 0,$$

as $n \to \infty$. Since $v_n(u) \sim u^{-1/\gamma}$ holds uniformly for $u \geq 1$, we get that

$$\sup_{u \geq 1} u^{1/(2\gamma)-\varepsilon} \left| \sqrt{k} \left(\frac{n}{k} \left(1 - \hat{F} \left(uU \left(\frac{n}{k} \right) \right) \right) - v_n(u) \right) - W_n(v_n(u)) \right| \xrightarrow{P} 0, \tag{3.5}$$

as $n \to \infty$. Comparing (3.5) with (3.2), to prove the lemma, it is only necessary to verify that, as $n \to \infty$,

$$\sup_{u \geq 1} u^{1/(2\gamma)-\varepsilon} \sqrt{k} \left| v_n(u) - u^{-1/\gamma} \right| \to 0,$$
$$\sup_{u \geq 1} u^{1/(2\gamma)-\varepsilon} \left| W_n \left(u^{-1/\gamma} \right) - W_n \left(v_n(u) \right) \right| \xrightarrow{P} 0.$$

The first relation follows from (3.3) and the limit relation $\lim_{n \to \infty} \sqrt{k} A(n/k) = 0$. The second relation can be proved by applying the Lévy's modulus of continuity theorem in a similar way as in the proof of Theorem 2.1, but simpler. □

Remark 3.3 *Although here the relation* (3.2) *is established only for* $\gamma > 0$, *one may obtain, following a similar proof, a corresponding result for general* $\gamma \in \mathbb{R}$. *The expression is the same as that in Theorem 5.1.2 in de Haan and Ferreira (2006). However, the result obtained here is based on non-stationary observations.*

Applying Proposition 3.2, we obtain the following theorem which gives the asymptotic normality of the Hill estimators with non-stationary observations under the second order conditions.

Theorem 3.4 *Let* $X_1, \cdots, X_n$ *be independent random variables following distribution functions* $F_{n,1}, \cdots, F_{n,n}$. *Suppose there exists a distribution function* F *such that the conditions* (1.1) *and* (1.2) *hold uniformly for all* $1 \leq i \leq n$. *Denote* $U(x) = \left(\frac{1}{1-F}\right)^{\leftarrow}(x)$. *Suppose that the second order conditions* (2.2) *and* (3.1) *hold for the same function* $A(t)$. *With a series of real number* $\{s_n\}_{n=1}^{\infty}$ *such that as* $n \to \infty$, $s_n \to +\infty$, $n(1-F(s_n)) \to +\infty$ *and* $\sqrt{n(1-F(s_n))}A(1/(1-F(s_n))) \to 0$, *define the Hill estimator with a deterministic threshold* s_n, $\hat{\gamma}_H(s_n)$, *as in* (1.4). *Moreover, with a series of integers* $k := k(n)$ *such that* $k \to +\infty$, $k/n \to 0$ *and* $\sqrt{k}A(n/k) \to 0$ *as* $n \to \infty$, *define the Hill estimator with a stochastic threshold,* $\hat{\gamma}_{H,k}$, *as in* (1.5). *We have that, as* $n \to \infty$,

$$\sqrt{n(1-F(s_n))}\left(\hat{\gamma}_H(s_n) - \gamma\right) \xrightarrow{d} N \quad \text{and} \quad \sqrt{k}\left(\hat{\gamma}_{H,k} - \gamma\right) \xrightarrow{d} N,$$

where N *is a normally distributed random variable with mean zero and variance* γ^2.

Remark 3.5 *The conditions on the deterministic threshold* s_n *and the intermediate sequence* $k(n)$ *are related to the same second order scale function* A *which appears in both* (2.2) *and* (3.1). *In fact, the two second order conditions can be allowed to have different scale functions* A_1 *and* A_2. *The theorem then holds with* $A := A_1 + A_2$.

Remark 3.6 *The consistency of the two Hill estimators is a direct consequence of Theorem 3.4. Nevertheless, similar to the discussion regarding the consistency of the tail empirical distribution function, the consistency is valid even without the second order conditions. This is consistent with the usual asymptotic results in extreme value statistics:*

for estimators of extreme value index, first order conditions imply consistency, second order conditions imply asymptotic normality.

Proof of Theorem 3.4 For the Hill estimator with deterministic threshold $\hat{\gamma}_H(s_n)$, denote $k = n(1 - F(s_n))$, or equivalently $s_n = U(n/k)$. It is not difficult to verify that k satisfies all the requirements on the $k(n)$ sequence used in the Hill estimator with stochastic threshold. The only difference is that k defined as $n(1 - F(s_n))$ may not be an integer as that used for the other estimator. However, the non-integer issue does not play a role the proof. Thus, without loss of generality, we use the same notation k for both estimators.

We start with writing the two Hill estimators as integrals of the empirical distribution function in the tail region. The Hill estimator with deterministic threshold can be rewritten as

$$\hat{\gamma}_H(s_n) = \frac{\int_{s_n}^{+\infty} \frac{1-\hat{F}(x)}{x} dx}{1 - \hat{F}(s_n)} = \frac{\int_1^{+\infty} 1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right) \frac{du}{u}}{1 - \hat{F}\left(U\left(\frac{n}{k}\right)\right)},$$

while the Hill estimator with stochastic threshold can be rewritten as

$$\hat{\gamma}_{H,k} = \frac{n}{k} \int_{X_{n,n-k}}^{+\infty} \frac{1 - \hat{F}(x)}{x} dx = \frac{n}{k} \int_{X_{n,n-k}/U(n/k)}^{+\infty} 1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right) \frac{du}{u}.$$

For the Hill estimator with deterministic threshold, by applying Proposition 3.2, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{n}{k}\int_1^{+\infty} 1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right)\frac{du}{u} - \int_1^{+\infty} u^{-1/\gamma}\frac{du}{u}\right) = \int_1^{+\infty} W_n\left(u^{-1/\gamma}\right)\frac{du}{u} + o_p(1).$$

Here, for the $o_p(1)$ term, we have used the fact that for $\varepsilon < 1/(2\gamma)$, $\lim_{n\to\infty} \int_1^{+\infty} u^{-1/(2\gamma)+\varepsilon} \frac{du}{u} = \frac{1}{1/(2\gamma)-\varepsilon} < +\infty$. With further simplification, we can write that as $n \to \infty$,

$$\sqrt{k}\left(\frac{n}{k}\int_1^{+\infty} 1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right)\frac{du}{u} - \gamma\right) = \gamma\int_0^1 W_n(t)\frac{dt}{t} + o_p(1). \tag{3.6}$$

Moreover, by taking $u = 1$ in (3.2), we get that, as $n \to \infty$,

$$\sqrt{k}\left(\frac{n}{k}\left(1 - \hat{F}\left(U\left(\frac{n}{k}\right)\right)\right) - 1\right) = W_n(1) + o_p(1).$$

This, together with (3.6), yields that, as $n \to \infty$,

$$\begin{aligned}
\sqrt{k}(\hat{\gamma}_H(s_n) - \gamma) &= \sqrt{k}\left(\frac{\frac{n}{k}\int_1^{+\infty} 1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right)\frac{du}{u}}{\frac{n}{k}\left(1 - \hat{F}\left(U\left(\frac{n}{k}\right)\right)\right)} - \gamma\right) \\
&= \sqrt{k}\left(\frac{\gamma + \frac{1}{\sqrt{k}}\gamma\int_0^1 W_n(t)\frac{dt}{t} + o_p(1/\sqrt{k})}{1 + \frac{1}{\sqrt{k}}W_n(1) + o_p(1/\sqrt{k})} - \gamma\right) \\
&= \gamma\left(\int_0^1 W_n(t)\frac{dt}{t} - W_n(1)\right) + o_p(1).
\end{aligned}$$

Denote $N \overset{d}{=} \gamma\int_0^1 W_n(t)\frac{dt}{t} - \gamma W_n(1)$. It is straightforward to verify that N follows a normal distribution with mean zero and variance γ^2.

Next we deal with the Hill estimator with stochastic threshold. By considering a finite range $u \in [1, T]$ for some $T > 1$ in the relation (3.2), we obtain the limit relation without a weight function as

$$\sup_{1 \le u \le T}\left|\sqrt{k}\left(\frac{n}{k}\left(1 - \hat{F}\left(uU\left(\frac{n}{k}\right)\right)\right) - u^{-1/\gamma}\right) - W_n\left(u^{-1/\gamma}\right)\right| \xrightarrow{P} 0,$$

as $n \to \infty$. With applying Vervaat's Lemma to the above relation, we get that as $n \to \infty$,

$$\sup_{\delta \le u \le 1}\left|\sqrt{k}\left(\frac{X_{n,n-\lfloor ku \rfloor}}{U(n/k)} - u^{-\gamma}\right) - \gamma u^{-\gamma-1}W_n(u)\right| \xrightarrow{P} 0, \tag{3.7}$$

where $\delta = T^{-1/\gamma} > 0$ is a constant. Taking $u = 1$ in (3.7) yields that

$$\sqrt{k}\left(\frac{X_{n,n-k}}{U(n/k)} - 1\right) - \gamma W_n(1) \xrightarrow{P} 0.$$

By combining this relation with (3.6), the rest of the proof follows the same lines as in Example 5.1.5 in de Haan and Ferreira (2006). The asymptotic limit has the same representation as that for the Hill estimator with deterministic threshold, i.e. as $n \to \infty$,

$$\sqrt{k}\left(\hat{\gamma}_{H,k} - \gamma\right) = \gamma \left(\int_0^1 W_n\left(t\right) \frac{dt}{t} - W_n(1) \right) + o_p(1).$$

Hence the theorem is proved for the Hill estimator with stochastic threshold as well. □

Corollary 3.7 *Under the same conditions as in Theorem 3.4, by taking* $s_n = U(n/k)$*, we have*

$$\sqrt{k}(\hat{\gamma}_H(U(n/k)) - \hat{\gamma}_{H,k}) \xrightarrow{P} 0.$$

as $n \to \infty$*.*

Corollary 3.7 implies that with a suitable choice of the thresholds, not only the limit distributions of the two types of Hill estimators are identical, but also the difference between them converges to zero in probability. A special case is the case of i.i.d. observations. Although seems intuitive, to our best knowledge, such a result is novel even for the i.i.d case.

Remark 3.8 *The asymptotic relation* (3.2) *is comparable with the weighted tail empirical process result for heavy-tailed* F *in the i.i.d. case; see, e.g. Theorem 5.1.4 in de Haan and Ferreira (2006). A result for the quantile process is given in* (3.7)*. However, the relation is established on a region where* u *is bounded away from zero. To have a result for the full region* $0 \leq u \leq 1$*, a proper weight function is necessary. This is left for future research.*

4 Conclusion and further extension

In this paper, we establish statistical tools for analyzing the tail region of distribution functions based on non-stationary observations. The non-stationarity refers to the case that observations are drawn from non-identical distributions. We show that by assuming comparable tail distributions, the weighted tail empirical process based on non-stationary observations has similar asymptotic behavior as that in the i.i.d case. This provides the fundamental tool for performing further extreme value analysis based on those observations. Assuming that the distribution functions belong to the domain of attraction with a positive tail index, we prove that Hill-type estimators are still valid estimators for the extreme value index of the underlying distribution functions and their asymptotic properties are the same as those for i.i.d. observations.

The two Hill estimators investigated in this paper are asymptotically unbiased due to the fact that $\sqrt{k}A(n/k) \to 0$ as $n \to \infty$. With a faster increasing k, i.e. as $n \to \infty$ $\sqrt{k}A(n/k) \to \lambda$, finite, the estimators are asymptotically biased as in the i.i.d. case. The asymptotic bias stems from two sources: the approximation of the tail of F by a Pareto distribution and the approximation of $1 - F_{n,i}(x)$ by $c_{n,i}(1 - F(x))$. The asymptotic bias stemming from the former is the same as that in the i.i.d. case, while that from the latter is due to the non-stationarity of the observations. The calculation of the exact form of the asymptotic bias requires a detailed specification of the second order condition (3.1), and is left for future research.

We remark that although we only show the asymptotic normality of the Hill-type estimators, similar results on other estimators of the extreme value index for more general range of γ can be established following the asymptotic behavior on the weighted tail empirical process. In principle, the proof for the Hill estimators is based on expressing the estimators as a functional transformation of the tail empirical process. Following the same logic, any estimator that can be written as a proper functional of the tail empirical process will preserve their asymptotic behaviors when the observations are non-stationary. Examples are the probability weighted moment estimator for $\gamma < 1/2$ (Hosking and Wallis (1987)) and the maximum likelihood estimator for $\gamma > -1/2$ (Smith (1985)).

Appendix

Proof of Theorem 2.1 We start with a few notations. Recall that

$$\mathbb{S}_n(t) = \frac{1}{\sqrt{k}} \left(\sum_{i=1}^{n} 1_{\eta_i < c_{n,i} tk/n} - tk \right).$$

Denote $\tilde{\mathbb{S}}_n(t) = \mathbb{S}_n(t)/q(t)$ and $\tilde{W}(t) = W(t)/q(t)$, where $W(t)$ is a standard Brownian motion on $[0, 2]$.

The theorem states that the process $\tilde{\mathbb{S}}_n(t)$ converges to a Gaussian process $\tilde{W}(t)$ weakly in the $(D[0, 2], \mathcal{D}, \| \cdot \|)$ space, where $\| \cdot \|$ denotes the supremum norm. The proof follows similar lines as in Shorack (1979), but emphasizes on the tail region of the empirical process. Since the proof is rather long, we split it into the following lemmas.

Lemma A.1 *For any $d \geq 1$ and $0 \leq t_1 < \cdots < t_d \leq 2$, the finite dimensional marginal $(\mathbb{S}_n(t_1), \cdots, \mathbb{S}_n(t_d))$ converges to $(W(t_1), \cdots, W(t_d))$ in distribution.*

Denote $\bar{\mathbb{S}}_n(t) = \mathbb{S}_n([kt]/k)$. Then $\bar{\mathbb{S}}_n(t)$ is a process in the $D-$space which has the two following properties.

Lemma A.2 *For any $q \in \mathbb{Q}$, define $\tilde{\bar{\mathbb{S}}}_n(t) := \bar{\mathbb{S}}_n(t)/q(t)$. Then $\left\{ \tilde{\bar{\mathbb{S}}}_n(t) : n \geq 1 \right\}$ is tight.*

Lemma A.3 *For any $q \in \mathbb{Q}$, as $n \to \infty$,*

$$\| \tilde{\mathbb{S}}_n(t) - \tilde{\bar{\mathbb{S}}}_n(t) \| \xrightarrow{P} 0.$$

For the notations, we remark that the sign $\bar{\cdot}$ gives a step version of the underlying process, while the sign $\tilde{\cdot}$ gives a weighted version weighing by the q function. For example, the process $\tilde{\bar{\mathbb{S}}}_n$ is thus the weighted version of the process $\bar{\mathbb{S}}_n$ that is a step version of the original $\mathbb{S}_n$ process.

Theorem 2.1 follows from the above three lemmas with the following logic. By considering a bounded q function, Lemma A.3 implies that as $n \to \infty$, $\| \mathbb{S}_n(t) - \bar{\mathbb{S}}_n(t) \| \xrightarrow{P} 0$. Together with Lemma A.1, we get that the process $\bar{\mathbb{S}}_n(t)$ converges to $W(t)$ in terms of any finite dimensional marginal distribution, which also implies the convergence of the process $\tilde{\tilde{\mathbb{S}}}_n(t)$ to $\tilde{W}(t)$ in terms of finite dimensional marginal distribution. Combining with the tightness of the process $\tilde{\tilde{\mathbb{S}}}_n(t)$ guaranteed by Lemma A.2, we get the weak convergence of the process $\tilde{\tilde{\mathbb{S}}}_n(t)$ to $\tilde{W}(t)$ in the $(D[0,2], \mathcal{D}, \|\cdot\|)$ space. Using Lemma A.3 again, we get that the process $\tilde{\mathbb{S}}_n(t)$ weakly converges to $\tilde{W}(t)$ in the $D-$space equipped with the supremum norm. Theorem 2.1 is thus proved. □

The remaining part of the appendix is devoted to prove Lemma A.1–A.3.

Proof of Lemma A.1 For any $0 \le t_1 < \cdots < t_d \le 2$ and real number $y_1, \cdots, y_d$, we check that as $n \to \infty$,

$$\mathbb{S}_n(y_1, \cdots, y_d; t_1, \cdots t_d) := \sum_{j=1}^{d} y_j \mathbb{S}_n(t_j) \xrightarrow{d} \sum_{j=1}^{d} y_j W(t_j).$$

Note that

$$\mathbb{S}_n(y_1, \cdots, y_d; t_1, \cdots t_d) = \sum_{i=1}^{n} (Y_{n,i} - EY_{n,i}),$$

where $Y_{n,i} = \frac{1}{\sqrt{k}} \sum_{j=1}^{d} y_j 1_{\eta_i < c_{n,i} k t_j / n}$. We calculate $V(Y_{n,i})$ as

$$V(Y_{n,i}) = \frac{1}{k} \sum_{j=1}^{d} \sum_{l=1}^{d} y_j y_l (c_{n,i} t_j \wedge t_l - (c_{n,i})^2 t_j t_l k/n).$$

Hence

$$v_n^2 := V(\mathbb{S}_n(y_1, \cdots, y_d; t_1, \cdots t_d)) = \sum_{j=1}^{d} \sum_{l=1}^{d} y_j y_l t_j \wedge t_l - \frac{k}{n^2} \sum_{i=1}^{n} (c_{n,i})^2 \sum_{j=1}^{d} \sum_{l=1}^{d} y_j y_l t_j t_l$$

Since $k/n \to 0$ as $n \to \infty$ and $\frac{1}{n} \sum_{i=1}^{n} c_{n,i}^2 \le b^2$ where b is the upper bound of all $c_{n,i}$, we get that

$$\lim_{n \to \infty} v_n^2 = \sum_{j=1}^{d} \sum_{l=1}^{d} y_j y_l t_j \wedge t_l = V(\sum_{j=1}^{d} y_j W(t_j)).$$

We check the Lyapunov condition that

$$\lim_{n \to \infty} \frac{1}{v_n^3} \sum_{i=1}^{n} E \left| Y_{n,i} - EY_{n,i} \right|^3 = 0.$$

Since

$$
\begin{aligned}
\sum_{i=1}^{n} E\left|Y_{n,i}-EY_{n,i}\right|^3 &\le \sum_{i=1}^{n} E(|Y_{n,i}|+|EY_{n,i}|)^3\\
&\le 8\sum_{i=1}^{n} E\left|Y_{n,i}\right|^3+(|EY_{n,i}|)^3\\
&=: 8(I_1+I_2),
\end{aligned}
$$

with verifying that $\lim_{n\to\infty} I_1 = \lim_{n\to\infty} I_2 = 0$, the Lyapunov condition holds. Firstly, from $EY_{n,i} = \frac{\sqrt{k}}{n} c_{n,i} \sum_{j=1}^{d} y_j t_j$, we get that

$$
0 \le \lim_{n\to\infty} I_2 = \lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{\sqrt{k}}{n} c_{n,i} \sum_{j=1}^{d} y_j t_j \right)^3 \le \lim_{n\to\infty} n \left(\frac{\sqrt{k}}{n} b \right)^3 \left(\sum_{j=1}^{d} y_j t_j \right)^3 = 0.
$$

Secondly, from Hölder inequality, we get that

$$
\begin{aligned}
EY_{n,i}^3 &= \left(\sqrt{\frac{1}{k}}\right)^3 E\left(\sum_{j=1}^{d} y_j 1_{\eta_i<c_{n,i}t_j k/n}\right)^3\\
&\le \left(\sqrt{\frac{1}{k}}\right)^3 E\left(\sum_{j=1}^{d} y_j^{3/2}\right)^2 \sum_{j=1}^{d} 1_{\eta_i<c_{n,i}t_j k/n}\\
&= \frac{c_{n,i}}{n}\frac{1}{\sqrt{k}}\left(\sum_{j=1}^{d} y_j^{3/2}\right)^2 \sum_{j=1}^{d} t_j,
\end{aligned}
$$

which implies that

$$
0 \le \lim_{n\to\infty} I_1 \le \lim_{n\to\infty} \frac{1}{\sqrt{k}} \left(\sum_{j=1}^{d} y_j^{3/2}\right)^2 \sum_{j=1}^{d} t_j = 0.
$$

With checking the Lyapunov condition, we can apply the Lyapunov Central Limit Theorem to obtain that $\frac{\mathbb{S}_n(y_1,\cdots,y_d;t_1,\cdots t_d)}{v_n} \xrightarrow{d} N(0,1)$ as $n\to\infty$, which is equivalent to $\mathbb{S}_n(y_1,\cdots,y_d;t_1,\cdots t_d) \xrightarrow{d} \sum_{j=1}^{d} y_j W(t_j)$. Hence, we conclude that any finite dimensional marginal distribution of $\mathbb{S}_n(t)$ converges to that of $W(t)$ on $t\in[0,2]$, as $n\to\infty$. $\Box$

Proof of Lemma A.2 From Theorem 15.6 in Billingsley (1968), the tightness follows from the following inequality. There exists a series of measures $\{\mu_n : n \geq 1\}$ a finite measure μ all defined on $[0, 2]$, such that for any $0 \leq t_1 \leq t \leq t_2 \leq 2$,

$$E\left(\tilde{\tilde{\mathbb{S}}}_n(t) - \tilde{\tilde{\mathbb{S}}}_n(t_1)\right)^2 \left(\tilde{\tilde{\mathbb{S}}}_n(t_2) - \tilde{\tilde{\mathbb{S}}}_n(t)\right)^2 \leq \mu_n(t_1, t]\mu_n(t, t_2], \tag{A.1}$$

and $\mu_n(0, t] \to \mu(0, t]$ as $n \to \infty$ for all $0 \leq t \leq 2$. To construct proper measures for the inequality (A.1), we first need some properties of the q function implied by the fact that $q \in \mathbb{Q}$. They are given in (1.15), (3.1) and (3.2) in Shorack (1979). For any $0 \leq s < t \leq 2$,

$$s\left(\frac{1}{q(s)} - \frac{1}{q(t)}\right)^2 \leq \frac{t-s}{q^2(t)} \leq \int_s^t \frac{1}{q^2(u)} du. \tag{A.2}$$

Moreover, as $t \to 0$

$$(\log t)\frac{t}{q^2(t)} \to 0. \tag{A.3}$$

Secondly, we need the following Lemma on the $\mathbb{S}_n(t)$ process.

Lemma A.4 *The process $\mathbb{S}_n(t)$ satisfies the following inequalities. For any $0 \leq t_1 < t < t_2 \leq 2$,*

$$E(\mathbb{S}_n(t) - \mathbb{S}_n(t_1))^2(\mathbb{S}_n(t_2) - \mathbb{S}_n(t))^2 \leq 3(t - t_1)(t_2 - t). \tag{A.4}$$

$$E(\mathbb{S}_n(t_2) - \mathbb{S}_n(t_1))^4 \leq \frac{t_2 - t_1}{k} + 3(t_2 - t_1)^2 \tag{A.5}$$

Proof of Lemma A.4 Write

$$\mathbb{S}_n(t) - \mathbb{S}_n(t_1) = \sum_{i=1}^n \xi_i(t_1, t),$$

where $\xi_i(t_1, t) := \frac{1_{\eta_i \in \frac{c_{n,i}k}{n}[t_1,t)} - E1_{\eta_i \in \frac{c_{n,i}k}{n}[t_1,t)}}{\sqrt{k}}$. From the independence among η_i, $(\xi_i(t_1, t), \xi_i(t, t_2))$ are independent random variables. Moreover $E\xi_i(t_1, t) = E\xi_i(t, t_2) = 0$.

Hence, we have that

$$
\begin{aligned}
&E(\mathbb{S}_n(t)-\mathbb{S}_n(t_1))^2(\mathbb{S}_n(t_2)-\mathbb{S}_n(t))^2\\
=&E\left(\sum_{i=1}^n \xi_i(t_1,t)\right)^2\left(\sum_{i=1}^n \xi_i(t,t_2)\right)^2\\
=&\sum_{i=1}^n E\xi_i(t_1,t)^2\xi_i(t,t_2)^2+\sum_{i\neq j}E(\xi_i(t,t_1)\xi_j(t_2,t))^2+2\sum_{i\neq j}E(\xi_i(t,t_1)\xi_i(t_2,t))(\xi_j(t,t_1)\xi_j(t_2,t))\\
=:&J_1+J_2+J_3.
\end{aligned}
$$

We deal with the three parts separately. Denote

$$
\begin{aligned}
p_{i,1}&:=E1_{\eta_i\in\frac{c_{n,i}k}{n}[t_1,t)}=(t-t_1)c_{n,i}\frac{k}{n}\\
p_{i,2}&:=E1_{\eta_i\in\frac{c_{n,i}k}{n}[t,t_2)}=(t_2-t)c_{n,i}\frac{k}{n}\\
p_{i,3}&:=1-p_{i,1}-p_{i,2}
\end{aligned}
$$

First, with J_1, we have that

$$
\begin{aligned}
J_1&=\sum_{i=1}^n E(\xi_i(t,t_1)\xi_i(t_2,t))^2\\
&=\frac{1}{k^2}\left(\sum_{i=1}^n p_{i,1}(1-p_{i,1})^2p_{i,2}^2+p_{i,2}p_{i,1}^2(1-p_{i,2})^2+p_{i,3}p_{i,1}^2p_{i,2}^2\right)\\
&\le\frac{1}{k^2}3\sum_{i=1}^n p_{i,1}p_{i,2}\\
&=\frac{3(t-t_1)(t_2-t)}{n^2}\sum_{i=1}^n c_{n,i}^2.
\end{aligned}
$$

Second, with J_2, we have that

$$\begin{aligned}
J_2 &= \sum_{i\neq j} E(\xi_i(t,t_1))^2 E(\xi_j(t_2,t))^2 \\
&= \frac{1}{k^2}\sum_{i\neq j} p_{i,1}(1-p_{i,1})p_{j,2}(1-p_{j,2}) \\
&\leq \frac{1}{k^2}\sum_{i\neq j} p_{i,1}p_{j,2} \\
&= \frac{(t-t_1)(t_2-t)}{n^2}\sum_{i\neq j} c_{n,i}c_{n,j}.
\end{aligned}$$

Lastly, with J_3, we have that

$$\begin{aligned}
J_3 &= 2\sum_{i\neq j} E(\xi_i(t,t_1)\xi_i(t_2,t))E(\xi_j(t,t_1)\xi_j(t_2,t)) \\
&= \frac{2}{k^2}\sum_{i\neq j} p_{i,1}p_{i,2}p_{j,1}p_{j,2} \\
&\leq \frac{2(t-t_1)(t_2-t)}{k^2}\sum_{i\neq j} p_{i,1}p_{j,2} \\
&= \frac{2(t-t_1)(t_2-t)}{n^2}\sum_{i\neq j} c_{n,i}c_{n,j}.
\end{aligned}$$

By aggregating J_1, J_2, J_3, we get that

$$\begin{aligned}
&E(\mathbb{S}_n(t)-\mathbb{S}_n(t_1))^2(\mathbb{S}_n(t_2)-\mathbb{S}_n(t))^2 \\
\leq&\frac{3(t-t_1)(t_2-t)}{n^2}\sum_{i=1}^n c_{n,i}^2 + \frac{3(t-t_1)(t_2-t)}{n^2}\sum_{i\neq j} c_{n,i}c_{n,j} \\
=&\frac{3(t-t_1)(t_2-t)}{n^2}\sum_{i=1}^n c_{n,i}\sum_{j=1}^n c_{n,j} \\
=&3(t-t_1)(t_2-t).
\end{aligned}$$

Hence the inequality (A.4) is proved. The proof for the inequality (A.5) follows similar steps and is given as follows.

$$
\begin{aligned}
& E(\mathbb{S}_n(t_2)-\mathbb{S}_n(t_1))^4 \\
=& E\left(\sum_{i=1}^n \xi_i(t_1,t_2)\right)^4 \\
=& \sum_{i=1}^n E\xi_i(t_1,t_2)^4 + 3\sum_{i\neq j} E\xi_i(t_1,t_2)^2 E\xi_j(t_1,t_2)^2 \\
\leq& \frac{1}{k^2}\left(\sum_{i=1}^n (p_{i,1}+p_{i,2}) + 3\sum_{i\neq j}(p_{i,1}+p_{i,2})(p_{j,1}+p_{j,2})\right) \\
=& \frac{t_2-t_1}{k} + 3(t_2-t_1)^2 \frac{\sum_{i\neq j} c_{n,i}c_{n,j}}{n^2} \\
\leq& \frac{t_2-t_1}{k} + 3(t_2-t_1)^2. \ \square
\end{aligned}
$$

Now we turn to prove the inequality (A.1). Notice that

$$
\begin{aligned}
\left(\frac{\bar{\mathbb{S}}_n(t)}{q(t)} - \frac{\bar{\mathbb{S}}_n(t_1)}{q(t_1)}\right)^2 &= \left(\frac{\bar{\mathbb{S}}_n(t)-\bar{\mathbb{S}}_n(t_1)}{q(t)} + \bar{\mathbb{S}}_n(t_1)\left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)\right)^2 \\
&\leq 2\left(\frac{(\bar{\mathbb{S}}_n(t)-\bar{\mathbb{S}}_n(t_1))^2}{q^2(t)} + (\bar{\mathbb{S}}_n(t_1))^2\left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2\right) \\
&:= 2(I_1+I_2).
\end{aligned}
$$

Similarly we have that

$$
\left(\frac{\bar{\mathbb{S}}_n(t_2)}{q(t_2)} - \frac{\bar{\mathbb{S}}_n(t)}{q(t)}\right)^2 \leq 2(I_3+I_4),
$$

with $I_3 := \frac{(\bar{\mathbb{S}}_n(t_2)-\bar{\mathbb{S}}_n(t))^2}{q^2(t_2)}$ and $I_4 := (\bar{\mathbb{S}}_n(t))^2\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2$. Thus,

$$
\begin{aligned}
E\left(\tilde{\bar{\mathbb{S}}}_n(t)-\tilde{\bar{\mathbb{S}}}_n(t_1)\right)^2\left(\tilde{\bar{\mathbb{S}}}_n(t_2)-\tilde{\bar{\mathbb{S}}}_n(t)\right)^2 &= E\left(\frac{\bar{\mathbb{S}}_n(t)}{q(t)} - \frac{\bar{\mathbb{S}}_n(t_1)}{q(t_1)}\right)^2\left(\frac{\bar{\mathbb{S}}_n(t_2)}{q(t_2)} - \frac{\bar{\mathbb{S}}_n(t)}{q(t)}\right)^2 \\
&\leq 4(EI_1I_3 + EI_1I_4 + EI_2I_3 + EI_2I_4).
\end{aligned}
$$

It is clear that if $t < 1/k$, $I_1 = I_2 = 0$, hence all four terms are zero. Thus, we only establish the inequalities for the four terms for $t \geq 1/k$. They hold automatically for the case $t < 1/k$.

Firstly, we look at EI_1I_3 which is given as

$$EI_1I_3 = \frac{1}{q^2(t)q^2(t_2)}E(\mathbb{S}_n([kt]/k) - \mathbb{S}_n([kt_1]/k))^2(\mathbb{S}_n([kt_2]/k) - \mathbb{S}_n([kt]/k))^2.$$

From Lemma A.4 and (A.2), we get that,

$$\begin{aligned}
EI_1I_3 &\le 3\frac{t-t_1+1/k}{q^2(t)}\frac{t_2-t+1/k}{q^2(t_2)} \\
&\le 3\left(\int_{t_1}^{t}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(t)}\right)\left(\int_{t}^{t_2}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(t_2)}\right) \\
&\le 3\left(\int_{t_1}^{t}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right)\left(\int_{t}^{t_2}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right).
\end{aligned}$$

The last step comes from $t \ge 1/k$.

Secondly, we deal with EI_1I_4. We have that

$$\begin{aligned}
EI_1I_4 &= \frac{1}{q^2(t)}\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 E(\bar{\mathbb{S}}_n(t) - \bar{\mathbb{S}}_n(t_1))^2(\bar{\mathbb{S}}_n(t))^2 \\
&\le \frac{1}{q^2(t)}\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 E(\bar{\mathbb{S}}_n(t) - \bar{\mathbb{S}}_n(t_1))^2 \cdot 2\left((\bar{\mathbb{S}}_n(t_1))^2 + (\bar{\mathbb{S}}_n(t) - \bar{\mathbb{S}}_n(t_1))^2\right) \\
&\le \frac{2}{q^2(t)}\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \left(E(\bar{\mathbb{S}}_n(t) - \bar{\mathbb{S}}_n(t_1))^2(\bar{\mathbb{S}}_n(t_1))^2 + E(\bar{\mathbb{S}}_n(t) - \bar{\mathbb{S}}_n(t_1))^4\right) \\
&\le \frac{2}{q^2(t)}\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \left(3(t-t_1+1/k)(t_1+1/k) + \frac{1}{k}(t-t_1+1/k) + 3(t-t_1+1/k)^2\right) \\
&\le \frac{2}{q^2(t)}\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \left(6(t-t_1+1/k)(t+1/k) + \frac{1}{k}(t-t_1+1/k)\right) \\
&= 12\frac{t-t_1+1/k}{q^2(t)}\cdot\left(t+\frac{7}{6k}\right)\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \\
&\le 12\left(\int_{t_1}^{t}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right)\left(\int_{t}^{t_2}\frac{1}{q^2(u)}du + \frac{7}{6k}\frac{1}{q^2(1/k)}\right).
\end{aligned}$$

Thirdly, the way we handle EI_2I_3 is similar to that for EI_1I_3, but simpler:

$$\begin{aligned}
EI_2I_3 &\le 3(t_1+1/k)\left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2\frac{t_2-t+1/k}{q^2(t_2)} \\
&\le 3\left(\int_{t_1}^{t}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right)\left(\int_{t}^{t_2}\frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right).
\end{aligned}$$

Lastly, we deal with EI_2I_4, which is essentially the same as the way we handle EI_1I4.

$$\begin{aligned}
EI_2I_4 &= \left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2 \left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 E(\bar{\mathbb{S}}_n(t_1))^2(\bar{\mathbb{S}}_n(t))^2 \\
&\leq \left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2 \left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \cdot 2\left(E(\bar{\mathbb{S}}_n(t_1))^2(\bar{\mathbb{S}}_n(t_1) - \bar{\mathbb{S}}_n(t))^2 + E(\bar{\mathbb{S}}_n(t_1))^4\right) \\
&\leq \left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2 \left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \cdot 2\left(3(t_1 + 1/k)(t - t_1 + 1/k) + \frac{t_1}{k} + 3(t_1 + 1/k)^2\right) \\
&\leq \left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2 \left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \cdot 2\left(6(t_1 + 1/k)t + \frac{t_1 + 1/k}{k}\right) \\
&= 12(t_1 + 1/k)\left(\frac{1}{q(t)} - \frac{1}{q(t_1)}\right)^2 \cdot \left(t + \frac{7}{6k}\right)\left(\frac{1}{q(t_2)} - \frac{1}{q(t)}\right)^2 \\
&\leq 12\left(\int_{t_1}^{t} \frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right)\left(\int_{t}^{t_2} \frac{1}{q^2(u)}du + \frac{7}{6k}\frac{1}{q^2(1/k)}\right).
\end{aligned}$$

Combining the four terms, we get that

$$\begin{aligned}
&E\left(\tilde{\bar{\mathbb{S}}}_n(t) - \tilde{\bar{\mathbb{S}}}_n(t_1)\right)^2 \left(\tilde{\bar{\mathbb{S}}}_n(t_2) - \tilde{\bar{\mathbb{S}}}_n(t)\right)^2 \\
\leq& 4(EI_1I_3 + EI_1I_4 + EI_2I_3 + EI_2I_4) \\
\leq& 120\left(\int_{t_1}^{t} \frac{1}{q^2(u)}du + \frac{1}{k}\frac{1}{q^2(1/k)}\right)\left(\int_{t}^{t_2} \frac{1}{q^2(u)}du + \frac{7}{6k}\frac{1}{q^2(1/k)}\right) \\
\leq& \left(11\left(\int_{t_1}^{t} \frac{1}{q^2(u)}du + \frac{7}{6k}\frac{1}{q^2(1/k)}\right)\right)\left(11\left(\int_{t}^{t_2} \frac{1}{q^2(u)}du + \frac{7}{6k}\frac{1}{q^2(1/k)}\right)\right).
\end{aligned}$$

By denoting $\mu_n(s,t] = 11\left(\int_s^t \frac{1}{q^2(u)}du + \frac{7}{6k}\frac{1}{q^2(1/k)}\right)$, we get that the inequality (A.1) holds. We only have to verify further that $\mu_n(0,t] \to \mu(0,t]$ for a finite measure μ. In fact, as $n \to \infty$, $k =\to +\infty$, which implies that $\frac{7}{6k}\frac{1}{q^2(1/k)} \to 0$ because of the relation (A.3). Therefore, $\mu_n(0,t] \to \mu(0,t] = \int_0^t \frac{1}{q^(u)}du$ as $n \to \infty$. The lemma is thus proved. □

Proof of Lemma A.3 The proof follows similar lines as in Shorack (1979). Write

$$\begin{aligned}
\sup_{t\in[0,2]}\left|\tilde{\mathbb{S}}_n(t) - \tilde{\bar{\mathbb{S}}}_n(t)\right| &\leq \max_{1\leq l\leq(k-1)} \sup_{t\in[l/k,(l+1)/k]} \frac{|\mathbb{S}_n(t) - \mathbb{S}_n(l/k)|}{q(l/k)} + \sup_{0\leq t\leq 1/k}\frac{|\mathbb{S}_n(t)|}{q(t)} \\
&=: \max_{1\leq l\leq(k-1)} Z_{n,l} + Z_{n,0}.
\end{aligned}$$

We first deal with $Z_{n,l}$ for $1 \leq l \leq (k-1)$. We have that

$$
\begin{aligned}
Z_{n,l} &= \frac{1}{\sqrt{k}} \sup_{t \in [l/k,(l+1)/k]} \frac{\left| \sum_{i=1}^{n} 1_{\frac{\eta_i}{c_{n,i}} \frac{n}{k} \in [l/k,t)} - (t - l/k)k \right|}{q(l/k)} \\
&\leq \frac{1}{\sqrt{k}} \sup_{t \in [l/k,(l+1)/k]} \frac{\sum_{i=1}^{n} 1_{\frac{\eta_i}{c_{n,i}} \frac{n}{k} \in [l/k,t)} + 1}{q(l/k)} \\
&\leq \frac{1}{\sqrt{k}} \frac{\sum_{i=1}^{n} 1_{\frac{\eta_i}{c_{n,i}} \frac{n}{k} \in [l/k,(l+1)/k)} + 1}{q(l/k)} \\
&\leq \frac{1}{\sqrt{k}} \frac{\left| \sum_{i=1}^{n} 1_{\frac{\eta_i}{c_{n,i}} \in [l/n,(l+1)/n)} - 1 \right| + 2}{q(l/k)} \\
&=: \frac{1}{\sqrt{k}} \frac{Y_{n,l} + 2}{q(l/k)}.
\end{aligned}
$$

The relation (A.3) implies that $\sqrt{k}q(l/k) \geq \sqrt{k}q(1/k) \to +\infty$ as $k \to \infty$. Hence for any $\varepsilon > 0$, there exists a N_ε such that for all $n > N_\varepsilon$, $\sqrt{k}q(1/k) > \max(\frac{1}{\varepsilon^3}, \frac{4}{\varepsilon})$. Thus, we get that for $n > N_\varepsilon$,

$$
\begin{aligned}
P(Z_{n,l} \geq \varepsilon) &\leq P\left(Y_{n,l} \geq \sqrt{k}q(l/k)\varepsilon - 2\right) \\
&\leq P\left(Y_{n,l} \geq \frac{\sqrt{k}q(l/k)\varepsilon}{2}\right) \\
&= P\left(|\mathbb{S}_n((l+1)/k) - \mathbb{S}_n(l/k)| \geq \frac{q(l/k)\varepsilon}{2}\right) \\
&\leq 16 \frac{E(\mathbb{S}_n((l+1)/k) - \mathbb{S}_n(l/k))^4}{(q(l/k)\varepsilon)^4} \\
&\leq 16 \frac{\frac{1}{k^2} + 3\frac{1}{k^2}}{(q(l/k)\varepsilon)^4} \\
&= \frac{64}{\varepsilon^4} \left(\frac{1}{kq^2(l/k)}\right)^2 \\
&\leq \frac{64}{\varepsilon^4} \left(\frac{1}{kq^2(l/k)}\right) (\varepsilon^3)^2 \\
&\leq \frac{64\varepsilon^2}{kq^2(l/k)}
\end{aligned}
$$

In the above derivation, we use the inequality (A.5). Thus, we can derive the tail probability on the maximum of all $Z_{n,l}$ as

$$P\left(\max_{1\le l\le (k-1)} Z_{n,l} \ge \varepsilon\right) \le \sum_{l=1}^{k-1} P(Z_{n,l}\ge\varepsilon) \le 64\varepsilon^2 \sum_{l=1}^{k-1} \frac{1}{kq^2(l/k)} \le 64\varepsilon^2 \int_0^1 \frac{1}{q^2(u)}du,$$

which implies that

$$\max_{1\le l\le (k-1)} Z_{n,l} \xrightarrow{P} 0, \quad \text{as} \ \ n\to\infty.$$

Lastly, we show that $Z_{n,0} \xrightarrow{P} 0$ as $n\to\infty$. Denote $m = k\log k$. Then $m/k\to\infty$ as $n\to\infty$. Moreover, from the relation (A.3), we get that as $n\to\infty$,

$$kq^2(1/m) = \frac{k\log m}{k\log k}\cdot\frac{mq^2(1/m)}{\log m}\to\infty.$$

Write

$$Z_{n,0} = \sup_{0\le t\le 1/m}\frac{|\mathbb{S}_n(t)|}{q(t)} + \sup_{1/m\le t\le 1/k}\frac{|\mathbb{S}_n(t)|}{q(t)} =: Z_{n,0}^{(1)} + Z_{n,0}^{(2)}.$$

Following similar lines as in the derivation on $Z_{n,l}$, by using the fact that $kq^2(1/m)\to\infty$, it can be proved that as $n\to\infty$, $Z_{n,0}^{(2)}\xrightarrow{P} 0$.

Finally, the following two facts ensure that as $n\to\infty$, $Z_{n,0}^{(1)}\xrightarrow{P} 0$. Firstly, on the set $A_n := \left\{\frac{\eta_i}{bk/n} > \frac{1}{m} \text{ for all } 1\le i\le n\right\}$, where b is the upper bound of all $c_{n,i}$, we have that

$$Z_{n,0}^{(1)} \le \frac{1}{\sqrt{k}}\sup_{0\le t\le 1/m}\frac{tk}{q(t)} \le \frac{\sqrt{k}}{\sqrt{m}}\sup_{0\le t\le 1/m}\frac{\sqrt{t}}{q(t)} \le \frac{\sqrt{k}}{m}\frac{1}{q(1/m)} \le \frac{1}{\sqrt{m}q(1/m)}\to 0,$$

as $n\to\infty$, in which we use the fact that $\frac{\sqrt{t}}{q(t)}$ is an increasing function on $[0,2]$. Secondly, the probability of the A_n set is calculated by

$$P(A_n) = \prod_{i=1}^n\left(1-\frac{bk/n}{m}\right) \le \exp\left(-\sum_{i=1}^n\frac{bk/n}{m}\right) = \exp\left(-\frac{bk}{m}\right)\to 1,$$

as $n\to\infty$. Thus the lemma is proved. □

References

P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, 1968.

M. Csörgo and L. Horváth. *Weighted approximations in probability and statistics.* John Wiley & Sons, 1993.

R. Davis and S. Resnick. Tail estimates motivated by extreme value theory. *The Annals of Statistics*, 12(4):1467–1487, 1984.

L. de Haan and A. Ferreira. *Extreme value theory: an introduction.* Springer Verlag, 2006.

H. Drees. Weighted approximations of tail processes for $\beta-$mixing random variables. *The Annals of Applied Probability*, 10:1274–1301, 2000.

J.H.J. Einmahl. Poisson and gaussian approximation of weighted local empirical processes. *Stochastic processes and their applications*, 70(1):31–58, 1997.

C.M. Goldie and R.L. Smith. Slow variation with remainder: Theory and applications. *The Quarterly Journal of Mathematics*, 38(1):45–71, 1987.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

J.R.M. Hosking and J.R. Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349, 1987.

T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 19 (3):1547–1569, 1991.

M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Verlag, 1983.

D. Mejzler. On the problem of the limit distribution for the maximal term of a variational series. *L'vov Politechn. Inst. Naučn. Zap.*, Ser. Fiz.-Mat. 38:90–109, 1956.

G.R. Shorack. The weighted empirical process of row independent random variables with arbitrary distribution functions. *Statistica Neerlandica*, 33(4):169–189, 1979.

G.R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. Wiley, New York, 1986.

R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.