Statistica Sinica Preprint No: SS-2024-0222	
Title	Distributed Inference for Tail Risks
Manuscript ID	SS-2024-0222
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0222
Complete List of Authors	Liujun Chen, Deyuan Li and Chen Zhou
Corresponding Authors	Chen Zhou
E-mails	zhou@ese.eur.nl
Notice: Accepted author version.	

Statistica Sinica

Distributed inference for tail risks

Liujun Chen[1], Deyuan Li[2], and Chen Zhou[3]

[1] *University of Science and Technology of China*

[2] *Fudan University and* [3] *Erasmus University Rotterdam*

Abstract: For measuring tail risk with scarce extreme events, extreme value analysis is often invoked as the statistical tool to extrapolate to the tail of a distribution. The presence of large datasets benefits tail risk analysis by providing more observations for conducting extreme value analysis. However, large datasets can be stored distributedly preventing the possibility of directly analyzing them. In this paper, we introduce a comprehensive set of tools for examining the asymptotic behavior of tail empirical and quantile processes in the setting where data is distributed across multiple sources, for instance, when data are stored on multiple machines. Utilizing these tools, one can establish the oracle property for most distributed estimators in extreme value statistics in a straightforward way. We provide various examples to demonstrate the practicality and value of our proposed toolkit.

Keywords: oracle property, tail empirical process, tail quantile process, KMT inequality

1. Introduction

Financial and climate risk management requires risk forecasting for rare but high-impact events, typically referred to as extreme events. Extreme value analysis, statistical methods for analyzing the tail of a distribution, is a useful tool for modeling and analyzing such extremes (de Haan and Ferreira, 2006). In this paper, we consider tail risk analysis using a large dataset that is distributedly stored at various locations.

While the availability of large datasets in general benefits statistical analysis, such as extreme value analysis, it also presents at least three practical challenges to implementing conventional statistical procedures. Firstly, a combined dataset might not be available to one end user due to privacy concerns such as analyzing large insurance claims across various insurance companies (Embrechts et al., 2013). Since insurance companies are contracted for protecting privacy of their customers, it is impossible to combine all claims from different insurers into one massive dataset. Secondly, the computation cost to analyze a massive dataset can be expensive when implementing statistical procedures involving an optimization algorithm, such as maximum likelihood or loss minimization. Thirdly, storage constraints can arise when dealing with massive datasets, for instance, when the size of a dataset exceeds a computer's memory. Another example is to analyze online stream data, where data become available in a sequential manner (Gama et al., 2013).

One solution to overcome these challenges is to handle the massive datasets in batches, sometimes referred to as "distributedly stored". Divide and Conquer (DC) algorithms are often invoked when data are distributedly stored in multiple machines. Assume that the N observations are distributedly stored in m machines with n observations in each machine with $N = nm$. More specifically, let $\mathcal{D}_j = \left\{X_1^{(j)}, \ldots, X_n^{(j)}\right\}$ denote the observations in machine j, with $j = 1, 2, \ldots, m$. Further denote $\mathcal{D} = \bigcup_{j=1}^m \mathcal{D}_j$ as the hypothetically combined sample, sometimes referred to as the *oracle sample*.

Let $\widehat{\theta}(\mathcal{D}_j)$ be an estimator for the parameter of interest based on data $\mathcal{D}_j$. Without prior knowledge on the structure of the estimator, a standard DC algorithm averages these local estimators to obtain a distributed estimate, $\widehat{\theta}^D := m^{-1} \sum_{j=1}^m \widehat{\theta}(\mathcal{D}_j)$. The DC algorithm has at least three advantages. Firstly, it preserves privacy. For example, insurance companies can share some statistical results provided that that other companies cannot infer client level data from the shared results. Moreover, thanks to the independent computation on each individual machine, the DC algorithm can significantly improve computational efficiency by utilizing parallel computing. Lastly, the DC algorithm can overcome the challenge of storage constraint by analyzing a massive dataset in batches.

The DC algorithm has been increasingly applied across various statistical procedures in recent years. For example, Li et al. (2013) investigated the use of the DC algorithm for kernel estimation, while Chang et al. (2017) explored its application to local smoothing estimators. In the context of high-dimensional data, Lee et al. (2017) employed the DC algorithm for the LASSO problem, and Lian and Fan (2018) applied it to support vector machines. Additionally, Fan et al. (2019) applied the DC algorithm to principal component analysis, and Volgushev et al. (2019) used it for quantile regression. We refer readers to Gao et al. (2022) for a comprehensive overview of the applications of the DC algorithm in statistics.

Let $\widehat{\theta}^{Oracle} := \widehat{\theta}(\mathcal{D})$ denote the *oracle estimator*, i.e., the same statistical procedure applied to the oracle sample $\mathcal{D}$. Theoretically, the DC algorithm can be applied to a given statistical procedure only if $\widehat{\theta}^{D}$ and $\widehat{\theta}^{Oracle}$ share the same asymptotic behavior, which we call *oracle property*.

The DC estimator in extreme value analysis may not achieve the oracle property. For example, considering a distribution with a finite endpoint, a natural estimator for the endpoint is the sample maxima. The oracle estimator $\widehat{\theta}^{Oracle}$ can be obtained by taking the maximum of the local estimators as $\widehat{\theta}^{Oracle} = \max_{1\le j\le m}\left(\widehat{\theta}(\mathcal{D}_j)\right)$, with $\widehat{\theta}(\mathcal{D}_j) = \max\left(X_1^{(j)}, \ldots, X_n^{(j)}\right)$. However, taking the maximum of all the local estimators is based on the prior knowledge of the structure of this estimator. Without such structural knowledge, the standard DC algorithm would result in an estimator $\widehat{\theta}^D = m^{-1}\sum_{j=1}^m \widehat{\theta}(\mathcal{D}_j)$, which may fail the oracle property.

The application of the DC algorithm in extreme value statistics has primarily focused on the Hill estimator (Hill, 1975) for the extreme value index $\gamma > 0$. Chen et al. (2022) proposed the distributed Hill estimator and study the asymptotic behavior of the distributed Hill estimator, demonstrating sufficient, sometimes also necessary, conditions under which the distributed Hill estimator possesses the oracle property. Daouia et al. (2024) extended this work by proving a stronger result: the difference between the distributed Hill estimator and the oracle Hill estimator diminishes faster than the speed of convergence of the oracle Hill estimator. However, the proofs in both papers rely heavily on the specific structure of the Hill estimator, and cannot be generalized to validate the oracle property of other estimators in extreme value statistics.

In this paper, we provide a set of tools to prove the oracle property for most estimators in extreme value analysis. Instead of focusing on a specific estimator $\widehat{\theta}$, our approach allows for proving the oracle property for most estimators based on the peak-over-threshold (POT) approach in a straightforward manner. We illustrate this by providing examples for the Hill estimator (Hill, 1975), the probability weighted moment (PWM) estimator (Hosking and Wallis, 1987), and the maximum likelihood estimator (MLE, Drees et al. (2004)).

To achieve this, we establish weighted approximations of tail empirical processes and tail quantile processes for the distributed subsamples jointly, with linking these approximations to that for the oracle sample. Observing that with equal subsample sizes across different machines, the tail empirical process for the oracle sample is the average of the tail empirical processes based on the distributed subsamples, it seems trivial that they can be approximated by the same asymptotic limits. However, to aggregate the tail empirical processes in different machines, we need to make sure that the approximation errors in different machines are uniformly negligible. We achieve this mathematically difficult result by invoking Komlós-Major-Tusnády type inequalities (see e.g. Komlós et al. (1975)). Linking the weighted approximation of the tail empirical process based on the oracle sample to those of the tail empirical processes on each machine is an important intermediate step towards establishing similar links between the corresponding tail quantile processes.

By contrast, when handling tail quantile processes, we cannot follow similar steps as for tail empirical processes. The main difference is that the average of the tail quantile processes based on distributed subsamples in different machines is not equal to the tail quantile process based on the oracle sample. Linking the approximations of the tail quantile processes based on the distributed subsamples to that based on the oracle sample poses an additional layer of technical difficulty, which we will handle in Section 4.

The rest of this paper is organized as follows. In Section 2, we review the extreme value analysis, tail empirical process, tail quantile process and DC algorithms. Section 3 shows the weighted approximations of the tail empirical processes based on the distributed subsamples in a joint manner and links that to the weighted approximation of tail empirical process based on the oracle sample. Section 4 shows the analogous result for the weighted approximations of the tail quantile processes. We provide various examples in Section 5 to show how these tools can be used to prove the oracle property of extreme value estimators such as the estimators of extreme value index, high quantile, tail probability and endpoint. Section 6 extends the theoretical results to the case of heterogeneous subsample sizes. A real data application is given in Section 7. A concluding remark is made in Section 8. The technical proofs are deferred to the Supplementary Material, along with a simulation study showing the performance of the distributed estimators for the extreme value index and the high quantile.

Throughout the paper, $a(t) \sim b(t)$ means that $a(t)/b(t) \to 1$ as $t \to \infty$; $a(t) \asymp b(t)$ means that both $|a(t)/b(t)|$ and $|b(t)/a(t)|$ are $O(1)$ as $t \to \infty$.

2. Background

2.1 Extreme Value Statistics

Let $X_1, \ldots, X_N$ be independently and identically distributed (i.i.d.) random variables with distribution function F, which is in the maximum domain of attraction of an extreme value distribution G_γ with index $\gamma \in \mathbb{R}$, i.e. there exist a positive function a and a real function b such that,

$$\lim_{N\to\infty} F^N \{a(N)x + b(N)\} = G_\gamma(x) := \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\},$$

for all $1+\gamma x > 0$. We denote this assumption as $F \in D(G_\gamma)$, where γ is the so called extreme value index. Extreme value statistics considers estimating the extreme value index γ, the functions a and b, as well as other practically relevant quantities such as high quantile of F. For established results in extreme value statistics, we refer interested readers to monographs such as de Haan and Ferreira (2006) and Resnick (2007).

Write $U = \{1/(1-F)\}^{\leftarrow}$, where $^{\leftarrow}$ denotes the left-continuous inverse function. Then the necessary and sufficient condition for $F \in D(G_\gamma)$ with $\gamma \in \mathbb{R}$ is

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{2.1}$$

for all $x > 0$. We further assume that U satisfies the second order condition, which quantifies the speed of convergence in (2.1) as follows: there exists an eventually positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho < 0$ such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right). \tag{2.2}$$

Although the limits in (2.1) and (2.2) appear unnecessarily specific, they are in fact the only possible non-degenerate limits of the expressions; see de Haan and Stadtmüller (1996). The second order condition (2.2) is satisfied by commonly used parametric distributions; see, e.g., Alves et al. (2007).

Under this condition, one can find suitable normalizing functions such that the convergence in (2.2) holds uniformly as follows, see Corollary 2.3.7 in de Haan and Ferreira (2006). There exists functions $a_0(t) \sim a(t)$, $A_0(t) \sim A(t)$ and $b_0(t)$ such that, for any $\varepsilon, \delta > 0$, there exists $t_0 = t_0(\varepsilon, \delta)$ such that, for all $t, tx \geq t_0$,

$$\left| \frac{\frac{U(tx)-b_0(t)}{a_0(t)} - \frac{x^\gamma-1}{\gamma}}{A_0(t)} - \Psi(x) \right| \leq \varepsilon x^{\gamma+\rho} \max\left(x^\delta, x^{-\delta}\right), \tag{2.3}$$

where

$$\Psi(x) := \begin{cases} \frac{x^{\gamma+\rho}}{\gamma+\rho}, & \gamma+\rho \neq 0, \\ \log x, & \gamma+\rho = 0. \end{cases}$$

For the details of the expression of a_0, b_0 and A_0, see Corollary 2.3.7 in de Haan and Ferreira (2006).

2.2 Tail Empirical Process and Tail Quantile Process

In classical extreme value statistics, two key tools for establishing asymptotic theories are the tail empirical process and the tail quantile process. Let $l = l(N)$ be an intermediate sequence such that as $N \to \infty$, $l \to \infty, l/N \to 0$. The tail empirical process is defined as

$$Y_{N,l}(x) = \frac{N}{l}\bar{F}_N\left\{a_0\left(\frac{N}{l}\right)x + b_0\left(\frac{N}{l}\right)\right\}, \quad x \in \mathbb{R},$$

where $\bar{F}_N := 1 - F_N$ and F_N denotes the empirical cumulative distribution function $F_N(x) := N^{-1}\sum_{i=1}^N I(X_i \leq x)$.

Under the second order condition (2.2) and $\sqrt{l}A(N/l) = O(1)$ as $N \to \infty$, Drees et al. (2006) showed that, under proper Skorokhod construction, there exists a sequence of Brownian motions $\{W_N^*\}_{N\geq 1}$, such that, for any

$v > 0$,

$$\sup_{x\in\mathbb{D}} \{z(x)\}^{v-1/2}\Big|\sqrt{l}\,\{Y_{N,l}(x) - z(x)\} - W_N^*\{z(x)\} - \sqrt{l}A_0(N/l)\,\{z(x)\}^{1+\gamma}\,\Psi\{1/z(x)\}\Big| = o_P(1), \tag{2.4}$$

where $z(x)$ and $\mathbb{D}$ are defined in Theorem 1 below. The approximation of the tail empirical process is a useful tool in a wider context. For example, Drees et al. (2006) proposed a test for the extreme value condition, de Haan and Ferreira (2006, Example 5.1.5) established the asymptotic normality of the Hill estimator, both by using this result.

Analogous to the tail empirical process, Drees (1998) showed a weighted approximation of the tail quantile process. The tail quantile process is defined as

$$Q_{N,l}(s) = \frac{X_{N-[ls],N} - b_0\left(\frac{N}{l}\right)}{a_0\left(\frac{N}{l}\right)}, \quad s \in [0,1],$$

where $X_{N,N} \geq \cdots \geq X_{1,N}$ are the order statistics of the sample $\{X_1,\ldots,X_N\}$. Here and thereafter, we use $[x]$ to denote the largest integer less than or equal to x. Assume the second order condition (2.2) and $\sqrt{l}A\,(N/l) = O(1)$ as $N\to\infty$, with the same Brownian motions $\{W_N^*\}_{N\geq 1}$ in (2.4), we have that, for any $v > 0$,

$$\sup_{1/l\leq s\leq 1} s^{v+1/2+\gamma}\left|\sqrt{l}\left(Q_{N,l}(s) - \frac{s^{-\gamma}-1}{\gamma}\right) - s^{-\gamma-1}W_N^*(s) - \sqrt{l}A_0\left(\frac{N}{l}\right)\Psi(s^{-1})\right| = o_P(1). \tag{2.5}$$

Note that the POT approach in extreme value statistics often uses high order statistics $X_{N,N}, \ldots, X_{N-l,N}$. Consequently, compared to the tail empirical process, the tail quantile process is more straightforward for proving asymptotic theory for estimators in extreme value statistics based on the POT approach. By writing such estimators as a functional of $Q_{N,l}(s)$ and using the weighted approximation of the tail quantile process, one can derive their asymptotic behavior.

3. Distributed Tail Empirical Process

The result (2.4) is based on the oracle sample. We intend to provide an analogous result for the tail empirical processes based on the distributed subsamples in a joint manner. Recall that the N observations are distributedly stored in m machines with n observations in each machine. We will extend our analysis to the case of heterogeneous subsample sizes in Section 6. The tail empirical process based on the observations $\left\{X_1^{(j)}, \ldots, X_n^{(j)}\right\}$ in machine j is defined as

$$Y_{n,k}^{(j)}(x) = \frac{n}{k}\bar{F}_n^{(j)}\left\{a_0\left(\frac{n}{k}\right)x + b_0\left(\frac{n}{k}\right)\right\}, \qquad j = 1, \ldots, m,$$

where $\bar{F}_n^{(j)} := 1 - F_n^{(j)}$ and $F_n^{(j)}$ denotes the empirical distribution function based on the observations in machine j. Here $k = k(N)$ is an intermediate sequence such that $k \to \infty$ and $k/n \to 0$, as $N \to \infty$.

We intend to relate the asymptotics of $Y_{N,l}(x)$ and $m^{-1}\sum_{j=1}^{m} Y_{n,k}^{(j)}(x)$ where $l = km$. Without causing any ambiguity, we use the simplified notation $Y_N(x)$ and $Y_n^{(j)}(x)$ for the tail empirical processes based on the oracle sample and the sample in machine j, respectively.

Throughout this paper, let m, n, k be sequences of integers such that, $m = m(N) \to \infty, n = n(N) \to \infty, k = k(N) \to \infty$ and $k/n \to 0$ as $N \to \infty$. We assume the following conditions for the sequences k and m:

(A1) $\sqrt{km}A(n/k) = O(1)$ as $N \to \infty$.

(A2) $\eta := \liminf_{N\to\infty} \log k/\log m - 1 > 0$.

(A3) $km(\log k)^2/n = O(1)$ as $N \to \infty$.

Remark 1. Note that $n/k = N/(km)$ and hence $\sqrt{km}A(n/k) = \sqrt{km}A(N/(km))$. Condition (A1) is a typical condition assumed in extreme value analysis to guarantee finite asymptotic bias in the oracle estimator. Condition (A2) states that, the number of machines (m) should be smaller than the number of observations used in each machine (k). Similar conditions are assumed in the literature of distributed inference for other statistical procedures, see e.g. Corollary 3.4 in Volgushev et al. (2019) and Theorem 4 in Zhu et al. (2021). Condition (A3) is an additional technical condition, which requires that the number of observations (n) in each machine is at a sufficiently high level for given k and m.

Remark 2. One example for k and m satisfying conditions (A1)-(A3) can be given as follows. Let $m \asymp n^a$ for some $0 \le a < -\frac{(-1)\vee\rho}{1-(-1)\vee\rho}$, where ρ is the second parameter in (2.2), and $k \asymp n^b$ for some

$$a < b < \min\left(1-a, \frac{-2\rho - a}{-2\rho+1}\right),$$

then conditions (A1)-(A3) hold with $\eta = b/a - 1 > 0$.

The following theorem shows the weighted approximations of the tail empirical processes based on the distributed subsamples in a joint manner.

Theorem 1. *Suppose that the distribution function F satisfies the second order condition (2.2) with $\gamma \in \mathbb{R}$ and $\rho < 0$. Let m, n, k be sequences of integers satisfying conditions (A1)-(A3) and $x_0 > -1/(\gamma \vee 0)$. Then under suitable Skorokhod construction, there exist m independent sequences of Brownian motions $\left\{W_n^{(j)}\right\}_{n\ge1}$, $j = 1, \ldots, m$, such that for any $v \in ((2+\eta)^{-1}, 2^{-1})$, as $N \to \infty$,*

$$\max_{1\le j\le m} \sup_{x\in\mathbb{D}} \{z(x)\}^{v-1/2} \Big| \sqrt{km}\left\{Y_n^{(j)}(x) - z(x)\right\} - \sqrt{m} W_n^{(j)}\{z(x)\}$$
$$- \sqrt{km} A_0(n/k) \{z(x)\}^{1+\gamma} \Psi\{1/z(x)\} \Big| = o_P(1),$$

where

$$z(x) = (1+\gamma x)^{-1/\gamma}, \quad \mathbb{D} = \left\{x : x_0 \le x < \frac{1}{(-\gamma)\vee 0}\right\}.$$

Moreover, as $N \to \infty$,

$$\sup_{x\in\mathbb{D}} \{z(x)\}^{v-1/2} \Big| \sqrt{km}\,\{Y_N(x) - z(x)\} - W_N\{z(x)\} - \sqrt{km}A_0(n/k)\,\{z(x)\}^{1+\gamma}\,\Psi\{1/z(x)\} \Big| = o_P(1),$$

where $W_N = m^{-1/2}\sum_{j=1}^m W_n^{(j)}$ *is a version of the Brownian motion* W_N^* *in* (2.4).

For $\gamma > 0$, a similar but simpler result is given as follows.

Theorem 2. *Suppose that the distribution function* F *satisfies the second order condition* (2.2) *with* $\gamma > 0$ *and* $\rho < 0$. *Let* m, n, k *be sequences of real numbers that satisfy conditions (A1)-(A3) and* $\tilde{x}_0 > 0$. *Then under suitable Skorokhod construction, there exist* m *independent sequences of Brownian motions* $\left\{W_n^{(j)}, n \geq 1\right\}, j = 1, \ldots, m$, *such that for any* $v \in ((2+\eta)^{-1}, 2^{-1})$, *as* $N \to \infty$,

$$\max_{1\le j\le m}\sup_{x\ge\tilde{x}_0} x^{(1/2-v)/\gamma}\Big|\sqrt{km}\left\{\frac{n}{k}\bar{F}_n^{(j)}\left(xU(n/k)\right) - x^{-1/\gamma}\right\} - \sqrt{m}W_n^{(j)}(x^{-1/\gamma}) - \sqrt{km}A_0\left(\frac{n}{k}\right)x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho}\Big| = o_P(1).$$

Moreover, as $N \to \infty$,

$$\sup_{x\ge\tilde{x}_0} x^{(1/2-v)/\gamma}\Big|\sqrt{km}\left\{\frac{n}{k}\bar{F}_N\left(xU(n/k)\right) - x^{-1/\gamma}\right\} - W_N(x^{-1/\gamma}) - \sqrt{km}A_0\left(\frac{n}{k}\right)x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho}\Big| = o_P(1),$$

where $W_N = m^{-1/2}\sum_{j=1}^m W_n^{(j)}$.

Theorems 1 and 2 show that the tail empirical process for the oracle sample $Y_{N,l}(x)$ and the average of the tail empirical processes $Y_{n,k}^{(j)}(x)$ across all machines can be approximated by the same Brownian motion. To prove these theorems, we need a fundamental inequality to bound the approximation error of the tail empirical process $Y_n^{(j)}(x)$ to the Gaussian process in machine j, which is of independent interest. Consider a positive sequence $t = t(N) \to 0$ as $N \to \infty$, satisfying

$$(n/k)^{-1/2} \log k/t = O(1), \tag{3.6}$$

$$k^{1/2} A_0(n/k)/t = O(1), \tag{3.7}$$

$$\text{for some } \tilde{\varepsilon} > 0, \{A_0(n/k)\}^{1/2-\tilde{\varepsilon}}/t = o(1). \tag{3.8}$$

Proposition 1. *Suppose that the distribution function F satisfies the second order condition (2.2) with $\gamma \in \mathbb{R}$ and $\rho < 0$. Let t be a sequence of real numbers satisfying conditions (3.6)-(3.8) and $x_0 > -1/(\gamma \vee 0)$. Then for sufficiently large n, under suitable Skorokhod construction, there exist m independent sequences of Brownian motions $\left\{W_n^{(j)}, n \geq 1\right\}, j = 1, \ldots, m$ and a constant $C_1 = C_1(v) > 0$ such that, for any $v \in (0, 1/2)$,*

$$P\left(\delta_n^{(j)} \geq t\right) \leq C_1 r^{-\frac{1}{1/2-v}},$$

where

$$\delta_n^{(j)} = \sup_{x\in\mathbb{D}} \{z(x)\}^{v-1/2} \Big| \sqrt{k} \left\{ Y_n^{(j)}(x) - z(x) \right\} - W_n^{(j)} \{z(x)\} - \sqrt{k} A_0(n/k) \{z(x)\}^{1+\gamma} \Psi \{1/z(x)\} \Big|,$$

and $r = r(t,k)$ *is defined by* $k^{-v} r \log r = t$.

Proposition 1 guarantees that the approximation errors $\delta_n^{(j)}, j = 1, \ldots, m$ are uniformly negligible, which is a key step to prove Theorems 1 and 2.

4. Distributed Tail Quantile Processes

Most estimators in extreme value statistics cannot be expressed as the functional of the tail empirical processes in a straightforward way, except the Hill estimator. Therefore, the two theorems in Section 3, focusing on the asymptotic expansion of the tail empirical processes jointly, cannot be directly applied to prove oracle properties of extreme value estimators, except for the Hill estimator. For the exceptional Hill estimator, we show, in the proof of Corollary 2 below, the usefulness of Theorem 2. By contrast, most estimators based on the POT approach can be written as functionals of tail quantile process. For that reason, we further investigate the distributed tail quantile processes.

Again, the result in (2.5) is based on the oracle sample. We intend to provide weighted approximations of the tail quantile processes based on the distributed subsamples in a joint manner. The tail quantile process based on the observations in machine j is defined as

$$Q_{n,k}^{(j)}(s) = \frac{X_{n-[ks],n}^{(j)} - b_0\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)}, \qquad j = 1, \ldots, m,$$

where $X_{n,n}^{(j)} \geq \cdots \geq X_{1,n}^{(j)}$ are the order statistics of the observations in machine j.

We aim at linking the asymptotics of $Q_{N,l}(s)$ and $m^{-1}\sum_{j=1}^{m} Q_{n,k}^{(j)}(s)$ where $l = km$. Again, without causing any ambiguity, we use the simplified notation $Q_N(s)$ and $Q_n^{(j)}(s)$ for the tail quantile process based on the oracle sample and the sample in machine j, respectively. Since the average of the tail quantile processes based on distributed subsamples in m machines is *not* equal to the tail quantile process of the oracle sample, we cannot follow similar steps as in Section 3. Instead, we achieve our goal by "inverting" the result for the tail empirical processes. More specifically, we intend to replace x in Theorem 1 by $Q_n^{(j)}(s)$ for $s \in [k^{-1+\delta}, 1]$.

The following theorem shows that, with the same sequences of Brownian motions defined in Theorem 1: $\left\{W_n^{(j)}\right\}_{n\geq 1}$, $j = 1, \ldots, m$, the approximation errors of the tail quantile processes are uniformly negligible for $1 \leq j \leq m$.

Theorem 3. *Assume the same conditions as in Theorem 1. Then for any*

$v \in ((2+\eta)^{-1}, 1/2)$ and $\delta \in (0,1)$, as $N \to \infty$,

$$\max_{1\le j\le m} \sup_{k^{-1+\delta}\le s\le 1} s^{v+1/2+\gamma}\Big|\sqrt{km}\left(Q_n^{(j)}(s) - \frac{s^{-\gamma}-1}{\gamma}\right) - \sqrt{m}s^{-\gamma-1}W_n^{(j)}(s) - \sqrt{km}A_0\left(\frac{n}{k}\right)\Psi(s^{-1})\Big| = o_P(1).$$

Moreover, as $N \to \infty$,

$$\sup_{k^{-1+\delta}\le s\le 1} s^{v+1/2+\gamma}\Big|\sqrt{km}\left(Q_N(s) - \frac{s^{-\gamma}-1}{\gamma}\right) - s^{-\gamma-1}W_N(s) - \sqrt{km}A_0\left(\frac{n}{k}\right)\Psi(s^{-1})\Big| = o_P(1).$$

Here, $\left\{W_n^{(j)}\right\}_{n\ge 1}, j = 1, \ldots, m$ are the same Brownian motions constructed as in Theorem 1 and $W_N = m^{-1/2}\sum_{j=1}^m W_n^{(j)}$. Consequently, $m^{-1}\sum_{j=1}^m Q_n^{(j)}(s)$ has the same asymptotic expansion as that for $Q_N(s)$, uniformly for $s \in [k^{-1+\delta}, 1]$.

For $\gamma > 0$, a similar but simpler result is given as follows.

Theorem 4. *Assume the same conditions as in Theorem 2. Then for any $v \in ((2+\eta)^{-1}, 1/2)$ and $\delta \in (0,1)$, as $N \to \infty$,*

$$\max_{1\le j\le m} \sup_{k^{-1+\delta}\le s\le 1} s^{v+1/2+\gamma}\Big|\sqrt{km}\left(\frac{X_{n-[ks],n}^{(j)}}{U(n/k)} - s^{-\gamma}\right) - \sqrt{m}\gamma s^{-\gamma-1}W_n^{(j)}(s) - \gamma\sqrt{km}A_0\left(\frac{n}{k}\right)s^{-\gamma}\frac{s^{-\rho}-1}{\rho}\Big| = o_P(1).$$

Moreover, as $N \to \infty$,

$$\sup_{k^{-1+\delta}\le s\le 1} s^{v+1/2+\gamma}\Big|\sqrt{km}\left(\frac{X_{N-[kms],N}}{U(n/k)} - s^{-\gamma}\right) - \gamma s^{-\gamma-1}W_N(s) - \gamma\sqrt{km}A_0\left(\frac{n}{k}\right)s^{-\gamma}\frac{s^{-\rho}-1}{\rho}\Big| = o_P(1).$$

Here, $\left\{W_n^{(j)}, n \geq 1\right\}, j = 1, \ldots, m$ *are the same Brownian motions constructed as in Theorem 1 and* $W_N = m^{-1/2} \sum_{j=1}^{m} W_n^{(j)}$.

Theorem 3 and 4 provide useful tools for establishing the oracle property of extreme value estimators based on the POT approach. For example, using Theorem 3, one can immediately show that, the distributed Pickands estimator (Pickands III, 1975) achieves the oracle property since the distributed Pickands estimator is a functional of the tail quantile processes $Q_n^{(j)}(s)$ at three points $s = 1, 1/2$ and $1/4$. We leave this to the readers.

The following corollary, which is a direct consequence of Theorem 4 with applying the Cramér's delta method, can be used for proving asymptotic theory of the distributed Hill estimator. Again, we leave such a proof to the readers. The final oracle property is the same as stated in Corrolary 2 below.

Corollary 1. *Assume the same conditions as in Theorem 2. By the Cramér's delta method, we can obtain that, as* $N \to \infty$,

$$\max_{1 \leq j \leq m} \sup_{k^{-1+\delta} \leq s \leq 1} s^{v+1/2} \Big| \sqrt{km} \left(\frac{\log X_{n-[ks],n}^{(j)} - \log U\left(\frac{n}{k}\right)}{\gamma} + \log s \right) - \sqrt{m} \gamma s^{-1} W_n^{(j)}(s) - \gamma \sqrt{km} A_0 \left(\frac{n}{k}\right) \frac{1}{\gamma} \frac{s^{-\rho} - 1}{\rho} \Big| = o_P(1).$$

Our ultimate goal is to develop a tool for establishing the asymptotic theories and oracle property for a broad range of estimators beyond the

Pickands estimator and the Hill estimator. However, Theorem 3 alone may not be sufficient to achieve this goal. We use the probability weighted moment (PWM) estimator as an example to explain the remaining issue.

The PWM estimator in machine j is defined as

$$\widehat{\gamma}_{PWM}^{(j)} := \frac{P_n^{(j)} - 4Q_n^{(j)}}{P_n^{(j)} - 2Q_n^{(j)}},$$

where

$$P_n^{(j)} := \frac{1}{k}\sum_{i=1}^{k} X_{n-i+1,n}^{(j)} - X_{n-k,n}^{(j)}, \quad Q_n^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \frac{i-1}{k}\left(X_{n-i+1,n}^{(j)} - X_{n-k,n}^{(j)}\right).$$

The distributed PWM estimator is defined as the average of the m estimates from each machine:

$$\widehat{\gamma}_{PWM}^{D} = \frac{1}{m}\sum_{j=1}^{m} \widehat{\gamma}_{PWM}^{(j)}.$$

To establish the asymptotic theory for $\widehat{\gamma}_{PWM}^{D}$, we need to handle the asymptotic expansions of $P_n^{(j)}$ and $Q_n^{(j)}$ for $j = 1, \ldots, m$ in a joint manner. For $s \in [0, 1]$, define

$$f_n^{(j)}(s) = Q_n^{(j)}(s) - \frac{s^{-\gamma} - 1}{\gamma} - \frac{1}{\sqrt{k}} s^{-\gamma-1} W_n^{(j)}(s) - A_0\left(\frac{n}{k}\right) \Psi(s^{-1}). \quad (4.9)$$

Then we can write $P_n^{(j)}$ as

$$
\begin{aligned}
&\frac{P_n^{(j)}}{a_0\left(n/k\right)}\\
&=\int_0^1 \frac{X_{n-[ks],n}^{(j)}-X_{n-k,n}^{(j)}}{a_0\left(\frac{n}{k}\right)}ds\\
&=\int_0^1 \frac{s^{-\gamma}-1}{\gamma}ds+\frac{\int_0^1\left\{s^{-\gamma-1}W_n^{(j)}(s)-W_n^{(j)}(1)\right\}ds}{\sqrt{k}}+A_0\left(\frac{n}{k}\right)\int_0^1\left\{\Psi(s^{-1})-\Psi(1)\right\}ds\\
&\quad+\int_{k^{-1+\delta}}^1\left\{f_n^{(j)}(s)-f_n^{(j)}(1)\right\}ds+\int_0^{k^{-1+\delta}}\left\{f_n^{(j)}(s)-f_n^{(j)}(1)\right\}ds\\
&=:I_1+I_2+I_3+I_4+I_5.
\end{aligned}
$$

The three terms I_1, I_2 and I_3 can be handled in a similar way as handling analogous terms in the oracle PWM estimator. The integral I_4 can be handled using Theorem 3. However, handling the last integral I_5 requires some additional tools to deal with the "corner" of the tail quantile processes. Similarly, for $Q_n^{(j)}$, we need to handle a different integral in the "corner": $\int_0^{k^{-1+\delta}} s\left\{f_n^{(j)}(s)-f_n^{(j)}(1)\right\}ds$. To complete the toolkit for our purpose, we provide a general result regrading the joint asymptotic behavior of weighted integrals of the tail quantile processes in the corner area $[0,k^{-1+\delta}]$.

Proposition 2. *Assume the same conditions as in Theorem 1. Assume that a function g defined on (0,1) satisfies $0<g(s)\le Cs^{\beta}$ with $\beta>\gamma-\frac{\eta}{2(1+\eta)}+\frac{1}{1+\eta}\gamma I\left\{\gamma>0\right\}$. Then, there exists a sufficiently small constant*

$\delta > 0$, such that, as $N \to \infty$,

$$\sqrt{m} \max_{1 \le j \le m} \int_0^{k^{-1+\delta}} g(s) \Big| \sqrt{k} \left(Q_n^{(j)}(s) - \frac{s^{-\gamma} - 1}{\gamma} \right) - s^{-\gamma-1} W_n^{(j)}(s) - \sqrt{k} A_0 \left(\frac{n}{k} \right) \Psi(s^{-1}) \Big| ds = o_P(1).$$

The oracle property of most extreme value estimators based on the POT approach, including the PWM estimator, can be established by applying Theorem 3 and Proposition 2 together.

5. Application

In this section, we provide examples illustrating how to apply our tools to establish asymptotic theories, particular the oracle property, for the distributed versions of various estimators in extreme value statistics. Specifically, we focus on the Hill estimator, PWM estimator and MLE for the extreme value index, and estimators for high quantiles, endpoints, and tail probabilities.

5.1 Distributed inference for the Hill estimator

In this subsection, we apply the approximations of the tail empirical processes based on the distributed subsamples to establish the oracle property of the distributed Hill estimator. The Hill estimator in machine j is defined as

$$\widehat{\gamma}_H^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n}^{(j)} - \log X_{n-k,n}^{(j)}, \quad j = 1, \ldots, m.$$

The distributed Hill estimator is defined as the average of the m estimates from each machine: $\widehat{\gamma}_H^D := m^{-1}\sum_{j=1}^{m} \widehat{\gamma}_H^{(j)}$. And the oracle Hill estimator using the top $l = km$ exceedance ratios is

$$\widehat{\gamma}_H^{Oracle} := \frac{1}{km}\sum_{i=1}^{km} \log X_{N-i+1,N} - \log X_{N-km,N}.$$

Corollary 2. *Suppose that the distribution function F satisfies the second order condition* (2.2) *with $\gamma > 0$ and $\rho < 0$. Let m, n, k be sequences of real numbers that satisfy conditions (A1)-(A3). Then, the distributed Hill estimator achieves the oracle property, i.e. $\sqrt{km}\left(\widehat{\gamma}_H^D - \widehat{\gamma}_H^{Oracle}\right) = o_P(1)$, as $N \to \infty$.*

Remark 3. Although Chen et al. (2022) and Daouia et al. (2024) both established the oracle property of the distributed Hill estimator, we provide a much shorter and more straightforward proof using the tools developed in Section 3. A different proof using the tools developed in Section 4 is also possible and even simpler, which we leave to the readers. Chen et al. (2022) only showed that the limiting distribution of the distributed Hill estimator coincides with that of the oracle Hill estimator, but does not investigate the difference between the two estimators. By contrast, using the tools developed in Section 3 and Section 4, we obtain a stronger result.

5.2 Distributed inference for the PWM estimator

In this subsection, we take the distributed PWM estimator as an example to show how to apply Theorem 3 and Proposition 2 to establish its oracle property. The oracle PWM estimator is defined as

$$\widehat{\gamma}_{PWM}^{Oracle} := \frac{P_N - 4Q_N}{P_N - 2Q_N},$$

where P_N and Q_N are counterparts of $P_n^{(j)}$ and $Q_n^{(j)}$ based on the oracle sample, respectively.

Corollary 3. *Suppose that the distribution function F satisfies the second order condition* (2.2) *with $\gamma < 1/2$ and $\rho < 0$. Assume that conditions (A1)-(A3) hold with $\eta > \max\left\{0, \frac{2\gamma}{1/2-\gamma}\right\}$. Then, the distributed PWM estimator achieves the oracle property, i.e., $\sqrt{km}\left(\widehat{\gamma}_{PWM}^{D} - \widehat{\gamma}_{PWM}^{Oracle}\right) = o_P(1)$ as $N \to \infty$.*

5.3 Distributed inference for the MLE

The MLE for the extreme value index and the scale parameter based on the sample on machine j $(\gamma_{mle}^{(j)}, \sigma_{mle}^{(j)})$, is defined as the solution of the following equations:

$$\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\gamma^2}\log\left(1+\frac{\gamma}{\sigma}\left(X^{(j)}_{n-i+1,n}-X^{(j)}_{n-k,n}\right)\right) - \left(\frac{1}{\gamma}+1\right)\frac{(1/\sigma)\left(X^{(j)}_{n-i+1,n}-X^{(j)}_{n-k,n}\right)}{1+(\gamma/\sigma)\left(X^{(j)}_{n-i+1,n}-X^{(j)}_{n-k,n}\right)}=0, \tag{5.10}$$

$$\sum_{i=1}^{k}\left(\frac{1}{\gamma}+1\right)\frac{(\gamma/\sigma)\left(X^{(j)}_{n-i+1,n}-X^{(j)}_{n-k,n}\right)}{1+(\gamma/\sigma)\left(X^{(j)}_{n-i+1,n}-X^{(j)}_{n-k,n}\right)}=k.$$

The distributed MLE for the extreme value index and the scale parameter are defined as

$$\widehat{\gamma}^{D}_{mle}=\frac{1}{m}\sum_{j=1}^{m}\widehat{\gamma}^{(j)}_{mle},\quad \widehat{\sigma}^{D}_{mle}=\frac{1}{m}\sum_{j=1}^{m}\widehat{\sigma}^{(j)}_{mle}.$$

The oracle MLE for the extreme value index and the scale parameter $(\widehat{\gamma}^{Oracle}_{mle},\widehat{\sigma}^{Oracle}_{mle})$ are defined in a similar way by using the oracle sample.

Corollary 4. *Suppose that the distribution function F satisfies the second order condition* (2.2) *with $\gamma>-1/2$ and $\rho<0$. Assume that conditions (A1)-(A3) hold with $\eta>\max\left(0,2\gamma,\frac{-2\gamma}{1+2\gamma}\right)$. Then, the distributed MLE for the extreme value index and the scale parameter achieve the oracle property, i.e., as $N\to\infty$,*

$$\sqrt{km}\left(\widehat{\gamma}^{D}_{mle}-\widehat{\gamma}^{Oracle}_{mle}\right)=o_P(1).$$

$$\sqrt{km}\,\frac{\widehat{\sigma}^{D}_{mle}-\widehat{\sigma}^{Oracle}_{mle}}{a(n/k)}=o_P(1).$$

Solving the likelihood equations (5.10) involves an optimization algorithm. The computation cost can be high when implementing an optimization algorithm for the oracle sample. We provide a simulation study to compare the computation cost of the oracle MLE and the distributed MLE in the Supplementary Material.

5.4 Distributed inference for the high quantile, endpoint and tail probability

In this subsection, we show how to establish the oracle property of the estimators for the high quantile, endpoint and tail probability. In order to estimate these quantities, we need to estimate the extreme value index γ, the scale parameter $a(n/k)$ and the location parameter $b(n/k)$, see e.g. de Haan and Ferreira (2006, Chapter 4). We focus on the PWM estimators for γ and $a(n/k)$ as an example. Other estimators based on the POT approach can be treated in a similar way.

Based on the oracle sample, since $N/(km) = n/k$, one can estimate $a(n/k)$ and $b(n/k)$ as

$$\widehat{a}^{Oracle}\left(\frac{n}{k}\right) = \frac{2P_N Q_N}{P_N - 2Q_N}, \quad \widehat{b}^{Oracle}\left(\frac{n}{k}\right) = X_{N-[km],N},$$

see e.g. Hosking and Wallis (1987).

We apply the DC algorithm to estimate $a(n/k)$ and $b(n/k)$ based on distributed subsamples. Define the distributed scale estimator as

$$\widehat{a}^D\left(\frac{n}{k}\right) := \frac{1}{m}\sum_{j=1}^{m}\widehat{a}^{(j)}\left(\frac{n}{k}\right) = \frac{1}{m}\sum_{j=1}^{m}\frac{2P_n^{(j)}Q_n^{(j)}}{P_n^{(j)} - 2Q_n^{(j)}},$$

and the distributed location estimator as

$$\widehat{b}^D\left(\frac{n}{k}\right) = \frac{1}{m}\sum_{j=1}^{m} X_{n-k,n}^{(j)}.$$

Following similar steps as in proving the oracle property of $\widehat{\gamma}_{PWM}^D$, we can show that, as $N \to \infty$,

$$\sqrt{km}\,\frac{\widehat{a}^D\left(\frac{n}{k}\right) - \widehat{a}^{Oracle}\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} = o_P(1), \quad \sqrt{km}\,\frac{\widehat{b}^D\left(\frac{n}{k}\right) - \widehat{b}^{Oracle}\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} = o_P(1).$$

5.4.1 High quantile

Let $x(p_N) := U(1/p_N)$, where $p_N = o(k/n)$ as $N \to \infty$, be the quantile we want to estimate. In finance management, the high quantile is often referred to as value at risk, which is the most prominent risk measure. The detailed procedures of the distributed estimator for high quantile $x(p_N)$ are given as follows:

- On each machine j, we calculate $\widehat{\gamma}_{PWM}^{(j)}, \widehat{a}^{(j)}\left(\frac{n}{k}\right), X_{n-k,n}^{(j)}$ and transmit these values to the central machine.

- On the central machine, we take the average of the $\widehat{\gamma}_{PWM}^{(j)}, \widehat{a}^{(j)}\left(\frac{n}{k}\right), X_{n-k,n}^{(j)}$ statistics collected from the m machines to obtain $\widehat{\gamma}_{PWM}^D, \widehat{a}^D\left(\frac{n}{k}\right), \widehat{b}^D\left(\frac{n}{k}\right)$.

- On the central machine, we estimate $x(p_N)$ with $p_N \to 0$ by

$$\widehat{x}^D(p_N) = \widehat{b}^D\left(\frac{n}{k}\right) + \widehat{a}^D\left(\frac{n}{k}\right)\frac{\left(\frac{k}{np_N}\right)^{\widehat{\gamma}^D_{PWM}} - 1}{\widehat{\gamma}^D_{PWM}}. \tag{5.11}$$

The oracle high quantile estimator $\widehat{x}^{Oracle}(p_N)$ is defined in an analogous way as $\widehat{x}^D(p_N)$, with replacing $\widehat{\gamma}^D_{PWM}$, $\widehat{a}^D\left(\frac{n}{k}\right)$ and $\widehat{b}^D\left(\frac{n}{k}\right)$ by $\widehat{\gamma}^{Oracle}_{PWM}$, $\widehat{a}^{Oracle}\left(\frac{n}{k}\right)$ and $\widehat{b}^{Oracle}\left(\frac{n}{k}\right)$ in (5.11), respectively. Following the lines of the proof for the asymptotics of the oracle high quantile estimator, we can obtain the asymptotic normality of $\widehat{x}^D(p_N)$. Moreover, since $\widehat{\gamma}^D_{PWM}$, $\widehat{a}^D\left(\frac{n}{k}\right)$ and $\widehat{b}^D\left(\frac{n}{k}\right)$ possess the oracle property, $\widehat{x}^D(p_N)$ also achieves the oracle property due to applying the Cramér delta method. We present the result in the following corollary while omitting the proof.

Corollary 5. *Assume the same conditions as in Corollary 3. Suppose that $np_N = o(k)$ and $\log(Np_N) = o(\sqrt{km})$ as $N \to \infty$. Then, as $N \to \infty$,*

$$\sqrt{km}\frac{\widehat{x}^D(p_N) - \widehat{x}^{Oracle}(p_N)}{a\left(\frac{n}{k}\right)q_\gamma(d_N)} = o_P(1),$$

where $d_N = k/(np_N)$ and for $t > 1$, $q_\gamma(t) := \int_1^t s^{\gamma-1}\log s ds$.

5.4.2 Endpoint

Next, we consider the problem of estimating the endpoint of the distribution function F. Assume that $F \in D(G_\gamma)$ for some $\gamma < 0$. In this case the endpoint $x^* = \sup\{x : F(x) < 1\}$ is finite. The endpoint can be treated as a specific case of quantile by regarding p_N as 0. The distributed endpoint estimator can be defined as

$$\widehat{x}^{*,D} = \widehat{b}^D\left(\frac{n}{k}\right) - \frac{\widehat{a}^D\left(\frac{n}{k}\right)}{\widehat{\gamma}^D_{PWM}}.$$

The definition of the oracle endpoint estimator $\widehat{x}^{*,Oracle}$ is in an analogous way. Again, the distributed endpoint estimator achieves the oracle property as in the following corollary.

Corollary 6. *Assume the same conditions as in Corollary 3 and $\gamma < 0$. Then, as $N \to \infty$,*

$$\sqrt{km}\frac{\widehat{x}^{*,D} - \widehat{x}^{*,Oracle}}{a\left(\frac{n}{k}\right)} = o_P(1).$$

5.4.3 Tail probability

Lastly, we consider the dual problem of estimating the high quantile: given a large value of x_N, how to estimate $p(x_N) = 1 - F(x_N)$ under the distributed inference setup. The detailed procedures for estimating the tail probability are similar to that for estimating the high quantile, except that on the central machine, we estimate the tail probability $p(x_N)$ by

$$\widehat{p}^D(x_N) = \frac{k}{n}\left\{\max\left(0, 1 + \widehat{\gamma}^D_{PWM}\frac{x_N - \widehat{b}^D\left(\frac{n}{k}\right)}{\widehat{a}^D\left(\frac{n}{k}\right)}\right)\right\}^{-1/\widehat{\gamma}^D_{PWM}}.$$

The definition of the oracle tail probability estimator $\widehat{p}^{Oracle}(x_N)$ is in an analogous way. Note that $\widehat{p}^{Oracle}(x_N)$ is valid only for $\gamma > -1/2$ (cf. Remark 4.4.3 in de Haan and Ferreira (2006)). The oracle property of $\widehat{p}^{D}(x_N)$ is established in the following corollary.

Corollary 7. *Assume the same conditions as in Corollary 3 and* $\gamma \in (-1/2, 1/2)$. *Denote* $d_N = \frac{k}{np(x_N)}$ *and* $w_\gamma(t) = t^{-\gamma}\int_1^t s^{\gamma-1}\log s ds$ *for* $t > 0$. *Suppose that* $d_N \to \infty$ *and* $w_\gamma(d_N) = o(\sqrt{km})$ *as* $N \to \infty$, *then*

$$\frac{\sqrt{km}}{w_\gamma(d_N)} \frac{\widehat{p}^{D}(x_N) - \widehat{p}^{Oracle}(x_N)}{p(x_N)} = o_P(1).$$

6. Heterogeneous subsample sizes

In this section, we extend our results to the case of heterogeneous subsample sizes. We assume that the N observations are distributedly stored in m machines with n_j observations in machine j, $j = 1, \ldots, m$, i.e. $N = \sum_{j=1}^m n_j$. Moreover, we assume that all n_j, $j = 1, \ldots, m$ diverge in the same order. Mathematically, that is,

$$c_1 \le \min_{1\le j\le m} n_j m/N \le \max_{1\le j\le m} n_j m/N \le c_2 \tag{6.12}$$

for some positive constants c_1 and c_2 and all $N \ge 1$.

The tail empirical process based on the observations in machine j is now defined as

$$Y^{(j)}_{n_j,k_j}(x) = \frac{n_j}{k_j}\bar{F}^{(j)}_{n_j}\left\{a_0\left(\frac{n_j}{k_j}\right)x + b_0\left(\frac{n_j}{k_j}\right)\right\}, \quad j = 1,\dots,m,$$

where $\bar{F}^{(j)}_{n_j} := 1 - F^{(j)}_{n_j}$ and $F^{(j)}_{n_j}$ denotes the empirical distribution function based on the observations in machine j.

We choose $k_j, j = 1,\dots,m$ such that the ratios k_j/n_j are homogenous across all the m machines, i.e.,

$$k_1/n_1 = \cdots = k_m/n_m. \tag{6.13}$$

Denote $K = \sum_{j=1}^m k_j$, clearly $k_j/n_j = K/N, j = 1,\dots,m$. Then, we have that,

$$Y_{N,K}(x) = \sum_{j=1}^m \frac{n_j}{N} Y^{(j)}_{n_j,k_j}(x).$$

In other words, the oracle tail empirical process is a weighted average of the tail empirical processes based on the distributed subsamples, where the weights equal to the fraction of the observations on each machine.

Following similar steps as in proving Theorem 1, we obtain the following result.

Theorem 5. *Assume the same conditions as in Theorem 1 and conditions* (6.12) *and* (6.13)*. Then under proper Skorokhod construction, there exist m independent sequences of Brownian motions $\left\{W^{(j)}_{n_j}\right\}, j = 1,\dots,m$, such* *that for any* $v \in ((2+\eta)^{-1}, 1/2)$*, as* $N \to \infty$*,*

$$\max_{1\leq j\leq m} \sup_{x\in\mathbb{D}} \{z(x)\}^{v-1/2} \left| \sqrt{k_j m} \left\{ Y^{(j)}_{n_j,k_j}(x) - z(x) \right\} - \sqrt{m} W^{(j)}_{n_j} \{z(x)\} - \sqrt{k_j m} A_0(N/K) \{z(x)\}^{1+\gamma} \Psi \{1/z(x)\} \right| = o_P(1).$$

Moreover, as $N \to \infty$*,*

$$\sup_{x\in\mathbb{D}} \{z(x)\}^{v-1/2} \left| \sqrt{K} \{Y_{N,K}(x) - z(x)\} - W_N \{z(x)\} - \sqrt{K} A_0(N/K) \{z(x)\}^{1+\gamma} \Psi \{1/z(x)\} \right| = o_P(1),$$

where $W_N = \sum_{j=1}^m \sqrt{\frac{n_j}{N}} W^{(j)}_{n_j}$ *is also a Brownian motion.*

Similar results hold for the tail quantile processes as in Theorem 3. Eventually, we can re-establish the oracle property of the distributed estimators as follows. Suppose the oracle estimator is based on K top order statistics in the oracle sample. On each machine, we use the top $k_j = (n_j/N)K$ order statistics in the estimation. By taking a weighted average of the estimates from all machines using the weights $n_j/N, j = 1, \ldots, m$, to obtain the distributed estimator, the oracle property holds under the same conditions as in the homogenous case with similar proofs.

7. Real Data Application

We use a dataset containing car insurance claims in five states of United States: Iowa ($n_1 = 2601$), Kansas ($n_2 = 798$), Missouri ($n_3 = 3150$), Nebraska ($n_4 = 1703$), and Oklahoma ($n_5 = 882$). The total sample size is $N = 9134$. We work under a hypothesis scenario: each state cannot share its own data to others, but they are willing to share their statistical results. Then one can apply a DC algorithm for conducting extreme value statistics. Our target is to estimate the common extreme value index of the total claim amount. We consider the MLE instead of the Hill estimator considered by Chen et al. (2022) since we do not assume heavy tail at the first place.

Let $K = \sum_{j=1}^{5} k_j$ be the total number of exceedances used by the five states. As suggested in Section 6, we choose k_j as $k_j = [K \frac{n_j}{N}]$, and apply the MLE for each of the five states to obtain $\widehat{\gamma}_{mle}^{(j)}$, $j = 1, 2, \ldots, 5$. Then, we combine these five estimates to obtain the distributed MLE by

$$\widehat{\gamma}_{mle}^{D} = \sum_{j=1}^{5} \frac{n_j}{N} \widehat{\gamma}_{mle}^{(j)}.$$

The distributed MLE is plotted against different values of K in Figure 1, along with its 95% confidence interval. We also plot the oracle MLE in this figure. The distributed MLE is close to the oracle MLE for almost all levels of K and the oracle MLE always falls into the 95% confidence interval constructed based on the distributed MLE.

By choosing $K = 1000$, we obtain that the distributed MLE for the extreme value index is about 0.05. And we cannot reject the hypothesis

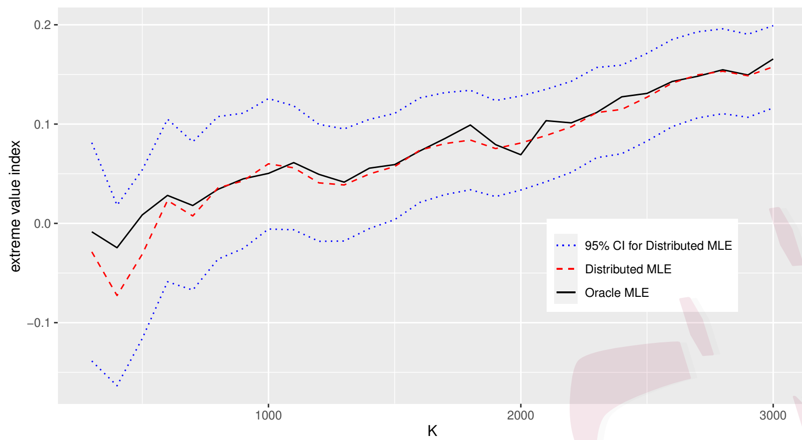

Figure 1: Car insurance data.

that the extreme value index is 0 under the 5% significance level for this choice of K. This result shows that the insurance claims may not be heavy tailed. In turn, the distributed Hill estimator adopted in Chen et al. (2022) may not be suitable for this application.

8. Discussion

In this paper, we investigate the problem of distributed inference in extreme value analysis when the oracle sample $\{X_1, X_2, \ldots, X_N\}$ are i.i.d.. In fact, the assumption that all the data are drawn from the same distribution can be relaxed. In real applications, observations from different machines may follow different distributions, but nevertheless share some common properties such as the extreme value index.

We assume that all observations are independent, but only observations on the same machine follow the same distribution. Denote the common distribution function of the observations in machine j as $F_{n,j}, j = 1, \ldots, m$. We assume that, there exists a continuous function F which satisfies the second order condition (2.2) with $\gamma > 0$. In addition, assume that the series of constants $\{c_{n,j}\}_{1 \le j \le m}$ satisfies that $0 < \underline{c} \le c_{n,j} \le \bar{c} < \infty$ for all $1 \le j \le m$ and $n \in \mathbb{N}$, and $A_1(t)$ is a positive regularly varying function with index $\tilde{\rho} < 0$ such that as $t \to \infty$,

$$\sup_{m \in \mathbb{N}} \max_{1 \le j \le m} \left| \frac{1 - F_{n,j}(t)}{1 - F(t)} - c_{n,j} \right| = O(A_1(t)).$$

By restricting that $\sqrt{km} A_1(n/k) \to 0$, Chen et al. (2022) gives a theoretical proof for the asymptotic theories of the distributed Hill estimator. Following similar steps, we can also handle tail empirical processes and tail quantile processes. The details are omitted.

Supplementary Materials

The Supplementary Material contains all the technical proofs and simulation studies.

Acknowledgements

The authors are grateful to the Editor, Associate Editor, and anonymous referees, whose suggestions led great improvement of this work. Liujun Chen's research was partially supported by the National Key R&D Program of China, No. 2024YFA1012200, and the National Natural Science Foundation of China grants 12301387 and 12471279. Deyuan Li's research was partially supported by the National Natural Science Foundation of China grants 11971115 and 12471279.

References

Alves, M. I. F., M. I. Gomes, L. de Haan, and C. Neves (2007). A note on second order conditions in extreme value theory: linking general and heavy tail conditions. *REVSTAT-Statistical Journal 5*(3), 285–304.

Chang, X., S.-B. Lin, and Y. Wang (2017). Divide and conquer local average regression. *Electronic Journal of Statistics 11*, 1326–1350.

Chen, L., D. Li, and C. Zhou (2022). Distributed inference for the extreme value index. *Biometrika 109*(1), 257–264.

Daouia, A., S. A. Padoan, and G. Stupfler (2024). Optimal weighted pooling for inference about the tail index and extreme quantiles. *Bernoulli 30*(2), 1287–1312.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* Springer Science

& Business Media.

de Haan, L. and U. Stadtmüller (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society 61*(3), 381–395.

Drees, H. (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics 25*(1), 187–210.

Drees, H., L. de Haan, and D. Li (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Journal of Statistical Planning and Inference 136*(10), 3498–3538.

Drees, H., A. Ferreira, and L. de Haan (2004). On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability 14*(3), 1179–1201.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2013). *Modelling Extremal Events: for Insurance and Finance*. Springer Science & Business Media.

Fan, J., D. Wang, K. Wang, and Z. Zhu (2019). Distributed estimation of principal eigenspaces. *Annals of Statistics 47*(6), 3009–3031.

Gama, J., R. Sebastiao, and P. P. Rodrigues (2013). On evaluating stream learning algorithms. *Machine Learning 90*, 317–346.

Gao, Y., W. Liu, H. Wang, X. Wang, Y. Yan, and R. Zhang (2022). A review of distributed statistical inference. *Statistical Theory and Related Fields 6*(2), 89–99.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution.

Annals of Statistics 3(5), 1163–1174.

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

Komlós, J., P. Major, and G. Tusnády (1975). An approximation of partial sums of independent rv'-s, and the sample df. i. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 32*(1), 111–131.

Lee, J. D., Q. Liu, Y. Sun, and J. E. Taylor (2017). Communication-efficient sparse regression. *Journal of Machine Learning Research 18*(5), 1–30.

Li, R., D. K. Lin, and B. Li (2013). Statistical inference in massive data sets. *Applied Stochastic Models in Business and Industry 29*(5), 399–409.

Lian, H. and Z. Fan (2018). Divide-and-conquer for debiased l_1-norm support vector machine in ultra-high dimensions. *Journal of Machine Learning Research 18*(182), 1–26.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*(1), 119–131.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Volgushev, S., S.-K. Chao, and G. Cheng (2019). Distributed inference for quantile regression processes. *Annals of Statistics 47*(3), 1634–1662.

Zhu, X., F. Li, and H. Wang (2021). Least squares approximation for a distributed system.

Journal of Computational and Graphical Statistics 30(4), 1004–1018.

International Institute of Finance, School of Management, University of Science and Technology of China

E-mail: ljchen22@ustc.edu.cn

School of Management, Fudan University

E-mail: deyuanli@fudan.edu.cn

Econometric Institute, Erasmus University Rotterdam

E-mail: zhou@ese.eur.nl