

THE CONVERGENCE RATE AND ASYMPTOTIC DISTRIBUTION OF THE BOOTSTRAP QUANTILE VARIANCE ESTIMATOR FOR IMPORTANCE SAMPLING

JINGCHEN LIU * ** AND

XUAN YANG,* *Columbia University*

Abstract

Importance sampling is a widely used variance reduction technique to compute sample quantiles such as value at risk. The variance of the weighted sample quantile estimator is usually a difficult quantity to compute. In this paper we present the exact convergence rate and asymptotic distributions of the bootstrap variance estimators for quantiles of weighted empirical distributions. Under regularity conditions, we show that the bootstrap variance estimator is asymptotically normal and has relative standard deviation of order $O(n^{-1/4})$.

Keywords: Weighted quantile; bootstrap; variance estimator; importance sampling

2010 Mathematics Subject Classification: Primary 62F40

Secondary 65C05

1. Introduction

In this paper we derive the asymptotic distributions of the bootstrap quantile variance estimators for weighted samples. Let F be a cumulative distribution function (CDF), let f be its density function, and let $\alpha_p = \inf\{x : F(x) \geq p\}$ be its p th quantile. It is well known that the asymptotic variance of the p th sample quantile is inversely proportional to $f(\alpha_p)$ (cf. [6]). When $f(\alpha_p)$ is close to 0 (e.g. p is close to 0 or 1), the sample quantile becomes very unstable since the ‘effective sample’ size is small. In the Monte Carlo scenario, one solution is to use importance sampling for variance reduction by distributing more samples around a neighborhood of the interesting quantile α_p . Such a technique has been widely employed in multiple disciplines. In portfolio risk management, the p th quantile of a portfolio’s total asset price is an important risk measure. This quantile is also known as the value at risk. Typically, the probability p in this context is very close to 0 (or 1). A partial literature list of using importance sampling to compute the value at risk includes [14], [23]–[25], [29], and [42]–[44]. In recent work Hult and Svensson [30] discussed efficient importance sampling for risk measure computation for heavy-tailed distributions. In the system stability assessment of engineering, the extreme quantile evaluation is of interest. In this context, the interesting probabilities are typically of a smaller order than those of the portfolio risk analysis.

Upon considering p to be close to 0 or 1, the computation of α_p can be viewed as the inverse problem of rare event simulation. The task of the latter topic is computing the tail probabilities $1 - F(b)$ when b tends to ∞ . Similar to the usage in quantile estimation, importance sampling is

Received 17 February 2011; revision received 15 December 2011.

* Postal address: Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA.

This research was supported in part by NSF grants CMMI-1069064 and SES-1123698, and the Institute of Education Sciences under grant R305D100017.

** Email address: jcliu@stat.columbia.edu

also a standard variance reduction technique for rare event simulation. The first to work on this topic was Siegmund [41], who not only presented an efficient importance sampling estimator but also defined a second-moment-based efficiency measure. We will later see that such a measure is also closely related to the asymptotic variance of the weighted quantiles. Such a connection allows us to adapt the efficient algorithms designed for rare event simulations to the computation of quantiles (cf. [26] and [30]). More recent works of rare event simulations for light-tailed distributions include [17], [19], and [39], and for heavy-tailed distributions include [3], [4], [8]–[11], [18], and [31]. There are also standard textbooks, such as [2] and [13].

Another field related to this line of work is survey sampling where unequal probability sampling and weighted samples are prevailing (cf. [32] and [36]). The weights are typically defined as the inverse of the inclusion probabilities.

The estimation of the distribution quantile is a classic topic. The almost-sure result of the sample quantile was established in [6]. The asymptotic distribution of the (unweighted) sample quantile can be found in a standard textbook, such as [15]. Estimation of the (unweighted) sample quantile variance via bootstrap was proposed in [5], [22], [37], [38], and [40]. There are also other kernel-based estimators (to estimate $f(\alpha_p)$) for such variances (cf. [21]).

There are several other works that are closely related to this work. The first is that of Hall and Martin [28], who derived the asymptotic distribution of the bootstrap quantile variance estimator for unweighted independent and identically distributed (i.i.d.) samples. Another is the work of Glynn [27], who derived the asymptotic distribution of weighted quantile estimators; see also [14] for a confidence interval construction. A more detailed discussion of these results is given in Section 2.2.

The asymptotic variance of the weighted sample quantile, as reported in [27], contains the density function $f(\alpha_p)$, whose evaluation typically consists of the computation of high-dimensional convolutions and is therefore usually not straightforward. In this paper we propose using the bootstrap method to compute/estimate the variance of such a weighted quantile. Bootstrap is a generic method that is easy to implement and does not consist of tuning parameters in contrast to the kernel-based methods for estimating $f(\alpha_p)$. In this paper we derive the convergence rate and asymptotic distribution of the bootstrap variance estimator for weighted quantiles. More specifically, the main contributions are to first provide conditions under which the quantiles of weighted samples have finite variances and develop their asymptotic approximations. Second, we derive the asymptotic distribution of the bootstrap estimators for such variances. Let n denote the sample size. Under regularity conditions (for instance, moment conditions and continuity conditions for the density functions), we show that the bootstrap variance estimator is asymptotically normal with a convergence rate of order $O(n^{-5/4})$. Given that the quantile variance decays at a rate of $O(n^{-1})$, the relative standard deviation of a bootstrap estimator is $O(n^{-1/4})$. Lastly, we present the asymptotic distribution of the bootstrap estimator for one particular case where $p \rightarrow 0$.

This work is technically challenging because many classic results of order statistics are not applicable. This is mainly caused by the variations introduced by the weights, which in the current context is the Radon–Nikodym derivative, and the fact that the weighted sample quantile does not map directly to the ordered statistics. In this paper we employ Edgeworth expansion combined with the strong approximation of empirical processes (see [33]) to derive the results.

This paper is organized as follows. In Section 2 we present our main results and summarize the related results in the literature. A numerical implementation is given in Section 3 to illustrate the performance of the bootstrap estimator. The proofs of the theorems are provided in Sections 4, 5, and 6.

2. Main results

2.1. Problem setting

Consider a probability space (Ω, \mathcal{F}, P) and a random variable X admitting CDF $F(x) = P(X \leq x)$ and density function

$$f(x) = F'(x)$$

for all $x \in \mathbb{R}$. Let α_p be its p th quantile, that is,

$$\alpha_p = \inf\{x : F(x) \geq p\}.$$

Consider a change of measure Q , under which X admits a CDF $G(x) = Q(X \leq x)$ and density

$$g(x) = G'(x).$$

Let

$$L(x) = \frac{f(x)}{g(x)},$$

and let X_1, \dots, X_n be i.i.d. copies of X under Q . Assume that P and Q are absolutely continuous with respect to each other. Then $E^Q L(X_i) = 1$. The corresponding weighted empirical CDF is

$$\hat{F}_X(x) = \frac{\sum_{i=1}^n L(X_i)I(X_i \leq x)}{\sum_{i=1}^n L(X_i)}, \tag{1}$$

where $I(\cdot)$ denotes the indicator function. A natural estimator of α_p is

$$\hat{\alpha}_p(\mathbf{X}) = \inf\{x \in \mathbb{R} : \hat{F}_X(x) \geq p\}. \tag{2}$$

Of interest in this paper is the variance of $\hat{\alpha}_p(\mathbf{X})$ under the sampling distribution of X_i , that is,

$$\sigma_n^2 = \text{var}^Q(\hat{\alpha}_p(\mathbf{X})). \tag{3}$$

The notation $E^Q(\cdot)$ and $\text{var}^Q(\cdot)$ are used to denote the expectation and variance under the measure Q .

Let Y_1, \dots, Y_n be i.i.d. bootstrap samples from the empirical distribution

$$\hat{G}(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x).$$

The bootstrap estimator for σ_n^2 in (3) is defined as

$$\hat{\sigma}_n^2 = \sum_{i=1}^n \hat{Q}(\hat{\alpha}_p(\mathbf{Y}) = X_i)(X_i - \hat{\alpha}_p(\mathbf{X}))^2, \tag{4}$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)$ and \hat{Q} is the measure induced by \hat{G} , that is, under \hat{Q} , Y_1, \dots, Y_n are i.i.d. with empirical distribution \hat{G} . Note that both \hat{G} and \hat{Q} depend on \mathbf{X} . To simplify the notation, we omit \mathbf{X} in the notation of \hat{Q} and \hat{G} .

Remark 1. There are multiple ways to form an estimate of F . One alternative to (1) is

$$\tilde{F}_X(x) = \frac{1}{n} \sum_{i=1}^n L(X_i)I(X_i \leq x). \tag{5}$$

The analysis of \tilde{F}_X is analogous to and simpler than that of (1). This is because the denominator is a constant. The weighted sample CDF in (5) depends only on samples below x . This is an important feature for the variance reduction of extreme quantile estimation when the change of measure Q is designed to be concentrated on the region below $F^{-1}(p)$. We will provide more detailed discussions later.

2.2. Related results

In this section we present two related results in the literature. First, Hall and Martin [28] established the asymptotic distribution of the bootstrap variance estimator for (unweighted) sample quantiles. In particular, they showed that if the density function $f(x)$ is Hölder continuous with index $\frac{1}{2} + \delta_0$ then

$$n^{5/4}(\hat{\sigma}_n^2 - \sigma_n^2) \Rightarrow N(0, 2\pi^{-1/2}[p(1-p)]^{3/2}f(\alpha_p)^{-4}) \tag{6}$$

as $n \rightarrow \infty$, where ‘ \Rightarrow ’ denotes weak convergence. This is consistent with the results in Theorem 2 below, setting $L(x) \equiv 1$. This paper can be viewed as a natural extension of [28], though the proof techniques are different.

In the context of importance sampling, as shown in [27], if $E^Q |L(x)|^3 < \infty$, the asymptotic distribution of a weighted quantile is

$$\sqrt{n}(\hat{\alpha}_p(X) - \alpha_p) \Rightarrow N\left(0, \frac{\text{var}^Q(W_p)}{f(\alpha_p)^2}\right) \tag{7}$$

as $n \rightarrow \infty$, where $W_p = L(X)(I(X < \alpha_p) - p)$. More general results in terms of weighted empirical processes are given in [30].

We now provide a brief discussion of the efficient quantile computation via importance sampling. The sample quantile admits a large variance when $f(\alpha_p)$ is small. One typical situation is that p is very close to 0 or 1. To fix ideas, we consider the case where p tends to 0. The asymptotic variance of the p th quantile of n i.i.d. samples is

$$\frac{1-p}{np} \frac{p^2}{f(\alpha_p)^2}.$$

Then, in order to obtain an estimate of an ε error with at least a $1 - \delta$ probability, the necessary number of i.i.d. samples is proportional to $p^{-1}p^2/f^2(\alpha_p)$, which grows to ∞ as $p \rightarrow 0$. Typically, the inverse of the hazard function, $p/f(\alpha_p)$, varies slowly as p tends to 0. For instance, $p/f(\alpha_p)$ is bounded if X is a light-tailed random variable and grows at most linearly in α_p for most heavy-tailed distributions (e.g. the regularly varying and log-normal distributions).

The asymptotic variance of the quantiles of \tilde{F}_X defined in (5) is

$$\frac{\text{var}^Q(L(X)I(X \leq \alpha_p))}{np^2} \frac{p^2}{f(\alpha_p)^2}.$$

There is a wealth of literature on the design of importance sampling algorithms, particularly those adapted to the context in which p is close to 0. A well-accepted efficiency measure is

precisely based on the relative variance $p^{-2} \text{var}^Q(L(X)I(X \leq \alpha_p))$ as $p \rightarrow 0$. More explicitly, the change of measure is called strongly efficient if $p^{-2} \text{var}^Q(L(X)I(X \leq \alpha_p))$ is bounded for arbitrarily small p . For recent developments in importance sampling algorithms in a rare event setting, see, e.g. [1], [7], [9], [12], and [17]. Therefore, the change of measure designed to estimate p can be adapted without much additional effort to the quantile estimation problem. For a more thorough discussion, see [14] and [30]. We will provide the analysis of one special case in Theorem 3 below.

2.3. The results for regular quantiles

In this subsection we provide an asymptotic approximation of σ_n^2 and the asymptotic distribution of $\hat{\sigma}_n^2$. We first list a set of conditions which we will refer to in the statements of our theorems.

(C1) There exists an $\alpha > 4$ such that

$$E^Q |L(X)|^\alpha < \infty.$$

(C2) There exists a $\beta > 3$ such that

$$E^Q |X|^\beta < \infty.$$

(C3) Assume that

$$\frac{\alpha}{3} > \frac{\beta + 2}{\beta - 3}.$$

(C4) There exists a $\delta_0 > 0$ such that the density functions $f(x)$ and $g(x)$ are Hölder continuous with index $\frac{1}{2} + \delta_0$ in a neighborhood of α_p , that is, there exists a constant c such that

$$|f(x) - f(y)| \leq c|x - y|^{1/2+\delta_0}, \quad |g(x) - g(y)| \leq c|x - y|^{1/2+\delta_0},$$

for all x and y in a neighborhood of α_p .

(C5) The measures P and Q are absolutely continuous with respect to each other. The likelihood ratio $L(x) \in (0, \infty)$ is Lipschitz continuous in a neighborhood of α_p .

(C6) Assume that $f(\alpha_p) > 0$.

Theorem 1. *Let F and G be the cumulative distribution functions of a random variable X under the probability measures P and Q , respectively. The distributions F and G have density functions $f(x) = F'(x)$ and $g(x) = G'(x)$. We assume that conditions (C1)–(C6) hold. Let*

$$W_p = L(X)I(X \leq \alpha_p) - pL(X),$$

and let $\hat{\alpha}_p(X)$ be as defined in (2). Then,

$$\sigma_n^2 := \text{var}^Q(\hat{\alpha}_p(X)) = \frac{\text{var}^Q(W_p)}{nf(\alpha_p)^2} + o(n^{-5/4}), \quad E^Q(\hat{\alpha}_p(X)) = \alpha_p + o(n^{-3/4}),$$

as $n \rightarrow \infty$.

Theorem 2. *Suppose that the conditions in Theorem 1 hold and that $L(X)$ has density under Q . Let $\hat{\sigma}_n^2$ be defined as in (4). Then, under Q ,*

$$n^{5/4}(\hat{\sigma}_n^2 - \sigma_n^2) \Rightarrow N(0, \tau_p^2) \tag{8}$$

as $n \rightarrow \infty$, where

$$\tau_p^2 = 2\pi^{-1/2}L(\alpha_p)f(\alpha_p)^{-4}(\text{var}^Q(W_p))^{3/2}.$$

Remark 2. In Theorem 1, we provided bounds on the errors of the asymptotic approximations for $E^Q(\hat{\alpha}_p(X))$ and σ_n^2 in order to assist the analysis of the bootstrap estimator. In particular, in order to approximate σ_n^2 with an accuracy of order $o(n^{-5/4})$, it is sufficient to approximate $E^Q(\hat{\alpha}_p(X))$ with an accuracy of order $o(n^{-5/8})$. Thus, Theorem 1 indicates that $\hat{\alpha}_p$ can be viewed as asymptotically unbiased. In addition, given that the bootstrap estimator has a convergence rate of $O_p(n^{-5/4})$, Theorem 1 suggests that, when computing the distribution of the bootstrap estimator, we can use the approximation of σ_n^2 to replace the true variance.

In Theorem 2, if we let $L(x) \equiv 1$, that is, $P = Q$, then $\hat{\alpha}_p$ is the regular quantile and the asymptotic distribution (8) recovers the result of [28] given in (6).

Remark 3. Note that the weak convergence in (7) requires weaker conditions than those in Theorems 1 and 2. The weak convergence does not require $\hat{\alpha}_p(X)$ to have a finite variance. In contrast, in order to apply the bootstrap variance estimator, we need to have the estimand well defined, that is, $\text{var}^Q(\hat{\alpha}_p(X)) < \infty$. Conditions (C1)–(C3) are imposed to ensure that $\hat{\alpha}_p(X)$ has a finite variance under Q .

The continuity assumptions on the density function f and the likelihood ratio function L (conditions (C4) and (C5)) are typically satisfied in practice. Condition (C6) is necessary for the quantile to have a variance of order $O(n^{-1})$.

2.4. Results for extreme quantile estimation

In this subsection we consider the particular case in which p tends to 0. The analysis in the context of extreme quantile estimation is sensitive to the underlying distribution and the choice of change of measure. Here, we only consider a stylized case, one of the first two cases considered in the rare event simulation literature. Let $X = \sum_{j=1}^m Z_j$, where the Z_j s are i.i.d. random variables with mean 0 and density function $h(z)$. The random variable X has density function $f(x)$ that is the m th convolution of $h(z)$. Note that both X and $f(x)$ depend on m . To simplify the notation, we omit the index m when there is no ambiguity. We further consider the exponential change of measure

$$Q(X \in dx) = e^{\theta x - m\varphi(\theta)} f(x) dx,$$

where $\varphi(\theta) = \log \int e^{\theta x} h(x) dx$. We say that φ is *steep* if, for every a , $\varphi(\theta) = a$ has a solution. For $\varepsilon > 0$, let $\alpha_p = -m\varepsilon$ be in the large deviations regime. We let θ be the solution to $\sup_{\theta'}(-\theta'\varepsilon - \varphi(\theta'))$ and $I = \sup_{\theta'}(-\theta'\varepsilon - \varphi(\theta'))$. Then, a well-known approximation of the tail probability is given by

$$P(X < -m\varepsilon) = \frac{c(\theta) + o(1)}{\sqrt{m}} e^{-mI}$$

as $m \rightarrow \infty$. The likelihood ratio is given by

$$LR(x) = e^{-\theta x + m\varphi(\theta)}.$$

We use the notation $LR(x)$ to distinguish it from the previous likelihood ratio $L(x)$.

Theorem 3. Suppose that $X = \sum_{j=1}^m Z_j$, where the Z_j s are i.i.d. mean-zero random variables with Lipschitz continuous density function. The log-moment-generating function $\varphi(\theta)$ is steep. For $\varepsilon > 0$, the equation

$$\varphi'(\theta) = -\varepsilon$$

has one solution denoted by θ . Let $\alpha_p = -m\varepsilon$, and let X_1, \dots, X_m be i.i.d. samples generated from the exponential change of measure

$$Q(X \in dx) = e^{\theta x - m\varphi(\theta)} f(x) dx.$$

Let

$$\tilde{F}_X(x) = \frac{1}{n} \sum_{i=1}^n LR(X_i)I(X_i \leq x), \quad \hat{\alpha}_p(X) = \inf(x \in \mathbb{R}: \tilde{F}_X(x) \geq p). \tag{9}$$

Let Y_1, \dots, Y_n be i.i.d. samples from the empirical measure \hat{Q} , and let $\hat{\sigma}_n^2$ be as defined in (4). If m (growing as a function of n and denoted by m_n) admits the limit $m_n^3/n \rightarrow c_* \in [0, +\infty)$ as $n \rightarrow \infty$ then

$$\frac{n^{5/4}}{\tilde{\tau}_p} (\hat{\sigma}_n^2 - \sigma_n^2) \Rightarrow N(0, 1)$$

as $n \rightarrow \infty$, where $\sigma_n^2 = \text{var}^Q(\hat{\alpha}_p(X))$,

$$\tilde{\tau}_p^2 = 2\pi^{-1/2} LR(\alpha_p) f(\alpha_p)^{-4} (\text{var}^Q(\bar{W}_p))^{3/2}, \quad \text{and} \quad \bar{W}_p = LR(X)I(X \leq \alpha_p).$$

Remark 4. Theorem 3 establishes the asymptotic distribution of the bootstrap variance estimator for a very stylized rare event simulation problem. For more general situations, further investigations are necessary; for example, in the heavy-tailed cases the likelihood ratios do not behave as well as those of the light-tailed cases even for strongly efficient estimators.

Remark 5. To simplify the notation, we drop the subscript n in the notation of m_n and write m whenever there is no ambiguity.

As m tends to ∞ , the term $\tilde{\tau}_p^2$ is no longer a constant. With the standard large deviations results (see, e.g. Lemma 6 below), we know that $\tilde{\tau}_p^2$ is of order $O(m^{5/4})$. Therefore, the convergence rate of $\hat{\sigma}_n^2$ is $O(m^{5/8}n^{-5/4})$. In addition, σ_n^2 is of order $O(m^{1/2}n^{-1})$. Thus, the relative convergence rate of $\hat{\sigma}_n^2$ is $O(m^{1/8}n^{-1/4})$. Choosing n so that $m^3 = O(n)$ is sufficient to estimate α_p with ε accuracy and σ_n^2 with ε -relative accuracy.

The empirical CDF in Theorem 3 is different from those in Theorems 1 and 2. We emphasize that it is necessary to use (9) to obtain the asymptotic results. This is mainly because the variance of $LR(X)$ grows exponentially fast as m tends to ∞ . Then, the normalizing constant of the empirical CDF in (1) is very unstable. In contrast, the empirical CDF in (9) depends only on the samples below x . Note that the change of measure is designed to reduce the variance of $LR(X)I(X \leq \alpha_p)$. Thus, the asymptotic results hold when n grows on the order of m^3 or faster.

3. A numerical example

In this section we provide a numerical example to illustrate the performance of the bootstrap variance estimator. In order to compare the bootstrap estimator with the asymptotic approximation in Theorem 1, we choose an example for which the marginal density $f(x)$ is in a closed form and α_p can be computed numerically. Consider the partial sum

$$X = \sum_{i=1}^m Z_i,$$

where the Z_i s are i.i.d. exponential random variables with rate 1. Then, the density function of X is

$$f(x) = \frac{x^{m-1}}{(m-1)!} e^{-x}.$$

TABLE 1: Comparison of variance estimators for fixed $m = 10$ and $n = 10\,000$. Here σ_n^2 is the quantile variance computed using crude Monte Carlo simulations, $\tilde{\sigma}_n^2$ is the asymptotic approximation of σ_n^2 in Theorem 1, and $\hat{\sigma}_n^2$ is the bootstrap estimate of σ_n^2 .

p	α_{1-p}	$\hat{\alpha}_{1-p}$	σ_n^2	$\tilde{\sigma}_n^2$	$\hat{\sigma}_n^2$
0.05	15.70	15.67	0.0032	0.0031	0.0027
0.04	16.16	16.13	0.0034	0.0033	0.0029
0.03	16.73	16.70	0.0037	0.0036	0.0032
0.02	17.51	17.47	0.0042	0.0041	0.0037
0.01	18.78	18.74	0.0054	0.0052	0.0047

TABLE 2: Comparison of variance estimators as $m \rightarrow \infty$ for $\alpha_p = 1.5m$ and $n = 10\,000$.

m	p	α_{1-p}	$\hat{\alpha}_{1-p}$	σ_n^2	$\tilde{\sigma}_n^2$	$\hat{\sigma}_n^2$
10	7.0×10^{-02}	15	14.95	1.1×10^{-03}	1.0×10^{-03}	1.2×10^{-03}
30	7.3×10^{-03}	45	44.95	2.2×10^{-03}	2.2×10^{-03}	1.7×10^{-03}
50	9.0×10^{-04}	75	74.98	3.0×10^{-03}	3.2×10^{-03}	2.4×10^{-03}
100	5.9×10^{-06}	150	149.99	4.8×10^{-03}	4.9×10^{-03}	5.0×10^{-03}

We are interested in computing the X_s $(1 - p)$ th quantile via an exponential change of measure, that is,

$$\frac{dQ_\theta}{dP} = \prod_{i=1}^m e^{\theta Z_i - \varphi(\theta)},$$

where $\varphi(\theta) = -\log(1 - \theta)$ for $\theta < 1$. We further choose $\theta = \arg \sup_{\theta'} (\theta' \alpha_p - m\varphi(\theta'))$.

We generate n i.i.d. replicates of (Z_1, \dots, Z_m) from Q_θ , that is, $(Z_1^{(k)}, \dots, Z_m^{(k)})$ for $k = 1, \dots, n$; then, we use $X_k = \sum_{i=1}^m Z_i^{(k)}$, $k = 1, \dots, n$, and the associated weights to form an empirical distribution and further $\hat{\alpha}_{1-p}(X)$. Let $\sigma_n^2 = \text{var}^{Q_\theta}(\hat{\alpha}_{1-p}(X))$, let $\tilde{\sigma}_n^2$ be the asymptotic approximation of σ_n^2 , and let $\hat{\sigma}_n^2$ be the bootstrap estimator of σ_n^2 . We use Monte Carlo simulations to compute both σ_n^2 and $\hat{\sigma}_n^2$ by generating independent replicates of $\hat{\alpha}_{1-p}(X)$ under Q and bootstrap samples under \hat{Q} , respectively.

We first consider the situation in which $m = 10$. In Table 1 we present the numerical results of estimators based on the empirical CDF in Theorem 2 with $n = 10\,000$. The column labeled σ_n^2 gives the variances of $\hat{\alpha}_{1-p}$ estimated using 1000 Monte Carlo simulations. In addition, we consider the case in which $\alpha_p = 1.5m$ and $m \rightarrow \infty$. In Table 2 we present the numerical results of the estimators given in Theorem 3 based on $n = 10\,000$ simulations.

4. Proof of Theorem 1

Throughout our discussion, we use the following notation for the asymptotic behavior. We say that $0 \leq g(b) = O(h(b))$ if $g(b) \leq ch(b)$ for some constant $c \in (0, \infty)$ and all $b \geq b_0 > 0$. Similarly, $g(b) = \Omega(h(b))$ if $g(b) \geq ch(b)$ for all $b \geq b_0 > 0$. We also write $g(b) = \Theta(h(b))$ if $g(b) = O(h(b))$ and $g(b) = \Omega(h(b))$. Finally, $g(b) = o(h(b))$ as $b \rightarrow \infty$ if $g(b)/h(b) \rightarrow 0$ as $b \rightarrow \infty$.

Before we present the proof of Theorem 1, we need a few useful lemmas.

Lemma 1. *Let X be a random variable with finite second moment. Then*

$$\begin{aligned} E X &= \int_{x>0} P(X > x) dx - \int_{x<0} P(X < x) dx, \\ E X^2 &= \int_{x>0} 2x P(X > x) dx - \int_{x<0} 2x P(X < x) dx. \end{aligned}$$

Lemma 2. *Let X_1, \dots, X_n be i.i.d. random variables with $E X_i = 0$ and $E |X_i|^\alpha < \infty$ for some $\alpha > 2$. For each $\varepsilon > 0$, there exists a constant κ depending on $\varepsilon, E X_i^2$, and $E |X_i|^\alpha$ such that*

$$E \left| \sum_{i=1}^n \frac{X_i}{\sqrt{n}} \right|^{\alpha-\varepsilon} \leq \kappa$$

for all $n > 0$.

The proofs of Lemmas 1 and 2 are elementary, and are thus omitted.

Lemma 3. *Let $h(x)$ be a nonnegative function. There exists $\zeta_0 > 0$ such that $h(x) \leq x^{\zeta_0}$ for all sufficiently large x . Then, for all $\zeta_1, \zeta_2, \lambda > 0$ such that $(\zeta_1 - 1)\lambda < \zeta_2$, we obtain*

$$\int_0^{n^\lambda} h(x) \Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) dx = \int_0^{n^\lambda} h(x) \Phi(-x) dx + o(n^{-\zeta_2})$$

as $n \rightarrow \infty$, where Φ is the CDF of a standard Gaussian distribution. In addition, we write $a_n(x) = o(x^{\zeta_1} n^{-\zeta_2})$ if $a_n(x)x^{-\zeta_1} n^{\zeta_2} \rightarrow 0$ as $n \rightarrow \infty$ uniformly for $x \in (\varepsilon, n^\lambda)$.

Proof. We first split the integral:

$$\begin{aligned} &\int_0^{n^\lambda} h(x) \Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) dx \\ &= \int_0^{(\log n)^2} h(x) \Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) dx + \int_{(\log n)^2}^{n^\lambda} h(x) \Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) dx. \end{aligned}$$

Note that we can bound the second term:

$$\int_{(\log n)^2}^{n^\lambda} h(x) \Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) dx \leq n^{\lambda(\zeta_0+1)} \Phi\left(-\frac{(\log n)^2}{2}\right) = o(n^{-\zeta_2}).$$

For the first term, note that, for all $0 \leq x \leq (\log n)^2$,

$$\Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) = (1 + o(x^{\zeta_1+1} n^{-\zeta_2})) \Phi(-x).$$

Then

$$\int_0^{(\log n)^2} h(x) \Phi(-x + o(x^{\zeta_1} n^{-\zeta_2})) dx = \int_0^{(\log n)^2} h(x) \Phi(-x) dx + o(n^{-\zeta_2}).$$

Therefore, the conclusion follows immediately.

Proof of Theorem 1. Let $\hat{\alpha}_p(X)$ be defined as in (2). To simplify the notation, we omit the index X and write $\hat{\alpha}_p(X)$ as $\hat{\alpha}_p$. We use Lemma 1 to compute the moments. In particular, we need to approximate the following probability:

$$\begin{aligned} Q(n^{1/2}(\hat{\alpha}_p - \alpha_p) > x) &= Q(\hat{F}_X(\alpha_p + xn^{-1/2}) < p) \\ &= Q\left(\sum_{i=1}^n L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) < 0\right). \end{aligned} \tag{10}$$

For some $\lambda \in (1/4(\alpha - 2), \frac{1}{8})$, we provide approximations for (10) in the following three cases: $0 < x \leq n^\lambda$, $n^\lambda \leq x \leq c\sqrt{n}$, and $x > \sqrt{n}$. The development for

$$Q(n^{1/2}(\hat{\alpha}_p - \alpha_p) < x)$$

in the region $x \leq 0$ is the same as that of the positive side.

Case 1: $0 < x \leq n^\lambda$. Let

$$W_{x,n,i} = L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) - F(\alpha_p + xn^{-1/2}) + p.$$

According to the Berry–Esseen bound (cf. [20]),

$$\begin{aligned} &Q\left(\sum_{i=1}^n L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) < 0\right) \\ &= Q\left(\frac{(1/n)\sum_{i=1}^n W_{x,n,i}}{\sqrt{\text{var}Q W_{x,n,1}/n}} < -\frac{F(\alpha_p + xn^{-1/2}) - p}{\sqrt{\text{var}Q W_{x,n,1}/n}}\right) \\ &= \Phi\left(\frac{-(F(\alpha_p + xn^{-1/2}) - p)}{\sqrt{\text{var}Q W_{x,n,1}/n}}\right) + D_1(x). \end{aligned} \tag{11}$$

There exists a constant κ_1 such that

$$|D_1(x)| \leq \frac{\kappa_1}{(\text{var}Q W_{x,n,1})^{3/2}} n^{-1/2}.$$

Case 2: $n^\lambda \leq x \leq c\sqrt{n}$. Thanks to Lemma 2, for each $\varepsilon > 0$, the $(\alpha - \varepsilon)$ th moment of $(1/\sqrt{n})\sum_{i=1}^n W_{x,n,i}$ is bounded. By Chebyshev’s inequality, we obtain

$$\begin{aligned} &Q\left(\sum_{i=1}^n L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) < 0\right) \\ &= Q\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n W_{x,n,i} \leq \sqrt{n}\left(p - F\left(\alpha_p + \frac{x}{\sqrt{n}}\right)\right)\right) \\ &\leq \kappa_1 \left(\frac{1}{\sqrt{n}(F(\alpha_p + xn^{-1/2}) - p)}\right)^{\alpha - \varepsilon} \\ &\leq \kappa_2 x^{-\alpha + \varepsilon}. \end{aligned}$$

Since $\lambda > 1/4(\alpha - 2)$, we choose ε small enough such that

$$\int_{n^\lambda}^{c\sqrt{n}} x Q(\hat{\alpha}_p - \alpha_p > xn^{-1/2}) dx = O(n^{-\lambda(\alpha - 2 - \varepsilon)}) = o(n^{-1/4}). \tag{12}$$

Case 3: $x > c\sqrt{n}$. Note that

$$Q\left(\sum_{i=1}^n L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) < 0\right)$$

is a nonincreasing function of x . Therefore, for all $x > c\sqrt{n}$, from case 2, we obtain

$$Q\left(\sum_{i=1}^n L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) < 0\right) \leq \kappa_2(c\sqrt{n})^{-\alpha+\varepsilon} = \kappa_3n^{-\alpha/2+\varepsilon/2}.$$

For $c\sqrt{n} < x \leq n^{\alpha/6-\varepsilon/6}$, we have

$$Q\left(\sum_{i=1}^n L(X_i)(I(X_i \leq \alpha_p + xn^{-1/2}) - p) < 0\right) \leq \kappa_3n^{-\alpha/2+\varepsilon/2} \leq \kappa_3x^{-3}.$$

In addition, note that, for all $x^{\beta-3} > n^{1+\beta/2}$,

$$\begin{aligned} Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) &\leq Q(\sup_i X_i > \alpha_p + xn^{-1/2}) \\ &= 1 - G^n(\alpha_p + xn^{-1/2}) \\ &\leq O(1)n^{1+\beta/2}x^{-\beta} \\ &= O(x^{-3}). \end{aligned}$$

Therefore, $Q(\hat{\alpha}_p > \alpha_p + x/\sqrt{n}) = O(x^{-3})$ on the region $\{c\sqrt{n} < x \leq n^{\alpha/6-\varepsilon/3}\} \cup \{x > n^{(\beta+2)/2(\beta-3)}\}$. Since $\alpha/3 > (\beta + 2)/(\beta - 3)$, we can choose ε small enough such that $x > n^{\alpha/6-\varepsilon/6}$ implies that $x^{\beta-3} > n^{1+\beta/2}$. Therefore, for all $x > c\sqrt{n}$, we obtain

$$Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) \leq x^{-3}$$

and

$$\int_{c\sqrt{n}}^\infty x Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) dx = O(n^{-1/2}). \tag{13}$$

Summarizing cases 2 and 3, more specifically (12) and (13), we obtain

$$\int_{n^\lambda}^\infty x Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) dx = o(n^{-1/4}).$$

Using the result in (11), we obtain

$$\begin{aligned} &\int_0^\infty x Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) dx \\ &= \int_0^{n^\lambda} x \left[\Phi\left(\frac{-(F(\alpha_p + xn^{-1/2}) - p)}{\sqrt{\text{var}^Q W_{x,n,1}/n}}\right) + O(n^{-1/2}) \right] dx + o(n^{-1/4}) \\ &= \int_0^{n^\lambda} x \Phi\left(\frac{-(F(\alpha_p + xn^{-1/2}) - p)}{\sqrt{\text{var}^Q W_{x,n,1}/n}}\right) dx + O(n^{2\lambda-1/2}) + o(n^{-1/4}). \end{aligned} \tag{14}$$

Given that $\lambda < \frac{1}{8}$, we have $O(n^{2\lambda-1/2}) = o(n^{-1/4})$. Thanks to condition (C4) and the fact that $\text{var}^Q(W_{x,n,1}) = (1 + O(xn^{-1/2})) \text{var}^Q(W_{0,n,1})$, we have

$$\frac{-(F(\alpha_p + xn^{-1/2}) - p)}{\sqrt{\text{var}^Q W_{x,n,1}/n}} = -\frac{xf(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n,1}}} + O(x^{3/2+\delta_0}n^{-1/4-\delta_0/2}).$$

Insert this approximation into (14). Together with the results from Lemma 3, we obtain

$$\int_0^{n^\lambda} x Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) dx = \int_0^{n^\lambda} x \Phi\left(-\frac{xf(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n,1}}}\right) dx + o(n^{-1/4}).$$

Therefore,

$$\begin{aligned} \int_0^\infty x Q(\hat{\alpha}_p - \alpha_p > x) dx &= \frac{1}{n} \int_0^\infty x Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) dx \\ &= \frac{1}{n} \int_0^\infty x \Phi\left(-\frac{xf(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n,1}}}\right) dx + o(n^{-5/4}) \\ &= \frac{\text{var}^Q W_{0,n,1}}{nf^2(\alpha_p)} \int_0^\infty x \Phi(-x) dx + o(n^{-5/4}) \\ &= \frac{\text{var}^Q W_{0,n,1}}{2nf^2(\alpha_p)} + o(n^{-5/4}). \end{aligned}$$

Similarly,

$$\begin{aligned} \int_0^\infty Q(\hat{\alpha}_p > \alpha_p + x) dx &= \frac{1}{\sqrt{n}} \int_0^\infty Q\left(\hat{\alpha}_p > \alpha_p + \frac{x}{\sqrt{n}}\right) dx \\ &= \frac{1}{\sqrt{n}} \int_0^\infty \Phi\left(-\frac{xf(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n,1}}}\right) dx + o(n^{-3/4}). \end{aligned}$$

For $Q(\hat{\alpha}_p < \alpha_p - x)$ and $x > 0$, the approximations are identical and are therefore omitted. Summarizing the results for $x > 0$ and $x \leq 0$ we obtain

$$\begin{aligned} E^Q(\hat{\alpha}_p - \alpha_p)^2 &= \int_0^\infty x Q(\hat{\alpha}_p > \alpha_p + x) dx + \int_0^\infty x Q(\hat{\alpha}_p < \alpha_p - x) dx \\ &= n^{-1} \left(\frac{\text{var}^Q W_{0,n,1}}{f^2(\alpha_p)} + o(n^{-1/4}) \right), \\ E^Q(\hat{\alpha}_p - \alpha_p) &= \int_0^\infty Q(\hat{\alpha}_p > \alpha_p + x) dx - \int_0^\infty Q(\hat{\alpha}_p < \alpha_p - x) dx \\ &= o(n^{-3/4}). \end{aligned}$$

5. Proof of Theorem 2

We first present a lemma that localizes the event. This lemma can be proven straightforwardly by standard results of empirical processes (cf. [33]–[35]) along with the strong law of large numbers and the central limit theorem. Therefore, we omit it. Let Y_1, \dots, Y_n be i.i.d. bootstrap samples, and let Y be a generic random variable equal in distribution to Y_i . Let \hat{Q} be the probability measure associated with the empirical distribution $\hat{G}(x) = (1/n) \sum_{i=1}^n I(X_i \leq x)$.

Lemma 4. Let C_n be the set in which the following events occur:

- (E1) $E^{\hat{Q}} |L(Y)|^\zeta < 2 E^Q |L(X)|^\zeta$, $\zeta = 2, 3, \alpha$; $E^{\hat{Q}} |L(Y)|^2 > \frac{1}{2} E^Q |L(X)|^2$; and $E^{\hat{Q}} |X|^\beta \leq 2 E^Q |X|^\beta$.
- (E2) Suppose that $\hat{\alpha}_p = X_{(r)}$. Then, assume that $|r/n - G(\alpha_p)| < n^{-1/2} \log n$ and $|\hat{\alpha}_p - \alpha_p| < n^{-1/2} \log n$.
- (E3) There exists $\delta \in (0, 1)$ such that, for all $1 < x < \sqrt{n}$,

$$\delta \leq \frac{\sum_{i=1}^n I(X_{(i)} \in (\hat{\alpha}_p, \hat{\alpha}_p + xn^{-1/2}])}{n Q(\alpha_p < X \leq \alpha_p + n^{-1/2}x)} \leq \delta^{-1}$$

and

$$\delta \leq \frac{\sum_{i=1}^n I(X_{(i)} \in (\hat{\alpha}_p - xn^{-1/2}, \hat{\alpha}_p])}{n Q(\alpha_p - n^{-1/2}x < X \leq \alpha_p)} \leq \delta^{-1}.$$

Then,

$$\lim_{n \rightarrow \infty} Q(C_n) = 1.$$

Lemma 5. Under conditions (C1) and (C5), let Y be a random variable with CDF \hat{G} . Then, for each $\lambda \in (0, \frac{1}{2})$,

$$\begin{aligned} & \sup_{|x| \leq cn^{\lambda-1/2}} |\text{var}^{\hat{Q}}[L(Y)(I(Y \leq \hat{\alpha}_p + x) - p)] - \text{var}^Q[L(X)(I(\alpha_p + x) - p)]| \\ & = O_p(n^{-1/2+\lambda}). \end{aligned}$$

Proof. Note that

$$\begin{aligned} & E^{\hat{Q}} L^2(Y)(I(Y \leq \hat{\alpha}_p + x) - p)^2 \\ & = \frac{1}{n} \sum_{i=1}^n L^2(X_i)(I(X_i \leq \hat{\alpha}_p + x) - p)^2 \\ & = \frac{1}{n} \sum_{i=1}^n L^2(X_i)(I(X_i \leq \alpha_p + x) - p)^2 \\ & \quad + L^2(\alpha_p) O_p\left(\frac{1}{n} \sum_{i=1}^n I(\min(\alpha_p, \hat{\alpha}_p) \leq X_i - x \leq \max(\alpha_p, \hat{\alpha}_p))\right). \end{aligned}$$

For the first term, by the central limit theorem, the continuity of $L(x)$, and Taylor’s expansion, we obtain

$$\begin{aligned} & \sup_{|x| \leq cn^{-1/2+\lambda}} \left| \frac{1}{n} \sum_{i=1}^n L^2(X_i)(I(X_i \leq \alpha_p + x) - p)^2 - E^Q(L^2(X)(I(X \leq \alpha_p + x) - p)^2) \right| \\ & = O_p(n^{-1/2+\lambda}). \end{aligned}$$

Thanks to the weak convergence of the empirical measure and $\hat{\alpha}_p - \alpha_p = O(n^{-1/2})$, we have

$$L^2(\alpha_p) O_p\left(\frac{1}{n} \sum_{i=1}^n I(\min(\alpha_p, \hat{\alpha}_p) \leq X_i - x \leq \max(\alpha_p, \hat{\alpha}_p))\right) = O_p(n^{-1/2}).$$

Therefore,

$$\begin{aligned} & \sup_{|x| \leq cn^{-1/2+\lambda}} |E^{\hat{Q}} L^2(Y)(I(Y \leq \hat{\alpha}_p + x) - p)^2 - E^Q(L^2(X)(I(X \leq \alpha_p + x) - p)^2)| \\ &= O_p(n^{-1/2+\lambda}). \end{aligned}$$

Using a very similar argument, we have

$$\begin{aligned} & \sup_{-cn^{-1/2} \leq x \leq cn^{-1/2}} |E^{\hat{Q}} L(Y)(I(Y \leq \hat{\alpha}_p + x) - p) - E^Q(L(X)(I(X \leq \alpha_p + x) - p))| \\ &= O_p(n^{-1/2+\lambda}). \end{aligned}$$

This completes the proof.

Proof of Theorem 2. Let $X_{(1)}, \dots, X_{(n)}$ be the order statistics of X_1, \dots, X_n in ascending order. Since we aim to prove weak convergence, it is sufficient to consider the case in which $X \in C_n$, as in Lemma 4. Throughout the proof, we assume that $X \in C_n$.

As in the proof of Theorem 1, we abbreviate $\hat{\alpha}_p(X)$ to $\hat{\alpha}_p$, but we keep the notation $\hat{\alpha}_p(Y)$ to differentiate between them. We use Lemma 1 to compute the second moment of $\hat{\alpha}_p(Y) - \hat{\alpha}_p$ under \hat{Q} , that is,

$$\hat{\sigma}_n^2 = \int_0^\infty x \hat{Q}(\hat{\alpha}_p(Y) > \hat{\alpha}_p + x) dx + \int_0^\infty x \hat{Q}(\hat{\alpha}_p(Y) < \hat{\alpha}_p - x) dx.$$

We first consider the case in which $x > 0$ and proceed using a derivation similar to that used in the proof of Theorem 1. Choose $\lambda \in (1/4(\alpha - 2), \frac{1}{8})$.

Case 1: $0 < x \leq n^\lambda$. Similarly to the proof of Theorem 1, by the Berry–Esseen bound, for all $x \in \mathbb{R}$,

$$\hat{Q}(n^{1/2}(\hat{\alpha}_p(Y) - \hat{\alpha}_p) > x) = \Phi\left(-\frac{\sum_{i=1}^n L(X_i)(I(X_i \leq \hat{\alpha}_p + xn^{-1/2}) - p)}{\sqrt{n \text{var}^{\hat{Q}} \tilde{W}_{x,n}}}\right) + D_2,$$

where

$$\tilde{W}_{x,n} = L(Y)\left(I\left(Y \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p\right) - \frac{1}{n} \sum_{i=1}^n L(X_i)\left(I\left(X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p\right)$$

and (thanks to (E1))

$$|D_2| \leq \frac{3 E^{\hat{Q}} |\tilde{W}_{x,n}|^3}{\sqrt{n}(\text{var}^{\hat{Q}} \tilde{W}_{x,n})^{3/2}} = O(n^{-1/2}).$$

In what follows, we further consider the cases in which $x > n^\lambda$. We will essentially follow cases 2 and 3 in the proof of Theorem 1.

Case 2: $n^\lambda \leq x \leq c\sqrt{n}$. Note that

$$\sum_{i=1}^n L(X_i)(I(X_i \leq \hat{\alpha}_p) - p) = O(1).$$

With exactly the same argument as used in case 2 of Theorem 1 and thanks to (E1), we obtain, for each $\varepsilon > 0$,

$$\begin{aligned} \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) - \hat{\alpha}_p > \frac{x}{\sqrt{n}}\right) &\leq \kappa \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n L(X_i) \left(I\left(X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p\right)\right)^{-\alpha+\varepsilon} \\ &= \kappa \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n L(X_i) I\left(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) + O\left(\frac{1}{\sqrt{n}}\right)\right)^{-\alpha+\varepsilon}. \end{aligned}$$

Furthermore, thanks to (E3), we have

$$\hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) - \hat{\alpha}_p > \frac{x}{\sqrt{n}}\right) = O(x^{-\alpha+\varepsilon}).$$

With sufficiently small ε , we have

$$\int_{n^\lambda}^{\sqrt{n}} x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) - \hat{\alpha}_p > \frac{x}{\sqrt{n}}\right) dx = O(n^{-\lambda(\alpha-\varepsilon-2)}) = o(n^{-1/4}).$$

Case 3: $x > c\sqrt{n}$. Note that

$$\hat{Q}\left(\sum_{i=1}^n L(Y_i) \left(I\left(Y_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p\right) < 0\right)$$

is a monotone nonincreasing function of x . Therefore, for all $x > c\sqrt{n}$, from case 2, we obtain

$$\hat{Q}\left(\sum_{i=1}^n L(Y_i) \left(I\left(Y_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p\right) < 0\right) \leq \kappa_3 n^{-\alpha/2+\varepsilon/2}.$$

For $x \leq n^{\alpha/6-\varepsilon/6}$, we obtain

$$\hat{Q}\left(\sum_{i=1}^n L(Y_i) \left(I\left(Y_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p\right) < 0\right) \leq \kappa_3 n^{-\alpha/2+\varepsilon/2} \leq \kappa_3 x^{-3}.$$

Thanks to condition (C3), with sufficiently small ε , $x > n^{\alpha/6-\varepsilon/6}$ implies that $x^{\beta-3} > n^{1+\beta/2}$. Therefore, because of (E1), for all $x > n^{\alpha/6-\varepsilon/6}$ (therefore, $x^{\beta-3} > n^{1+\beta/2}$),

$$\hat{Q}\left(\hat{\alpha}(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) \leq \hat{Q}\left(\sup_i Y_i > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) = O(1)n^{1+\beta/2}x^{-\beta} = O(x^{-3}).$$

Therefore, we have

$$\int_{c\sqrt{n}}^{\infty} x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) - \hat{\alpha}_p > \frac{x}{\sqrt{n}}\right) dx = O(n^{-1/2}).$$

From the results of cases 2 and 3, we obtain, for $X \in C_n$,

$$\int_{n^\lambda}^\infty x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) dx = o(n^{-1/4}). \tag{15}$$

Using exactly the same proof, we can show that

$$\int_{n^\lambda}^\infty x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) < \hat{\alpha}_p - \frac{x}{\sqrt{n}}\right) dx = o(n^{-1/4}). \tag{16}$$

Case 1 revisited. Cases 2 and 3 imply that the integral in the region where $|x| > n^\lambda$ can be ignored. In the region $0 \leq x \leq n^\lambda$, on the set C_n , for $\lambda < \frac{1}{8}$, we obtain

$$\begin{aligned} & \int_0^{n^\lambda} x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) dx \\ &= \int_0^{n^\lambda} x \left[\Phi\left(-\frac{\sum_{i=1}^n L(X_i)(I(X_i \leq \hat{\alpha}_p + xn^{-1/2}) - p)}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}}\right) + D_2 \right] dx \\ &= \int_0^{n^\lambda} x \Phi\left(-\frac{\sum_{i=1}^n L(X_i)(I(X_i \leq \hat{\alpha}_p + xn^{-1/2}) - p)}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}}\right) dx \\ &\quad + o(n^{-1/4}). \end{aligned} \tag{17}$$

We now take a closer look at the integrand. Note that

$$\begin{aligned} & \sum_{i=1}^n L(X_i) \left(I\left(X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p \right) \\ &= \sum_{i=1}^n L(X_i) (I(X_i \leq \hat{\alpha}_p) - p) + \sum_{i=1}^n L(X_i) I\left(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right). \end{aligned} \tag{18}$$

Suppose that $\hat{\alpha}_p = X_{(r)}$. Then

$$\sum_{i=1}^r L(X_{(i)}) \geq p \sum_{i=1}^n L(X_i) \quad \text{and} \quad \sum_{i=1}^{r-1} L(X_{(i)}) < p \sum_{i=1}^n L(X_i). \tag{19}$$

Therefore,

$$p \sum_{i=1}^n L(X_i) \leq \sum_{i=1}^n L(X_i) I(X_i \leq \hat{\alpha}_p) < L(\hat{\alpha}_p) + p \sum_{i=1}^n L(X_i).$$

Substituting this into (18) we obtain

$$\begin{aligned} & \sum_{i=1}^n L(X_i) \left(I\left(X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) - p \right) \\ &= O(L(\hat{\alpha}_p)) + \sum_{i=1}^n L(X_i) I\left(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right). \end{aligned} \tag{20}$$

In what follows, we study the dominating term in (17) via (20). For all $x \in (0, n^\lambda)$, thanks to (20), we obtain

$$\begin{aligned} & \Phi\left(-\frac{\sum_{i=1}^n L(X_i)(I(X_i \leq \hat{\alpha}_p + xn^{-1/2}) - p)}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}}\right) \\ &= \Phi\left(-\frac{\sum_{i=1}^n L(X_i)I(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + xn^{-1/2})}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}} + O(n^{-1/2})\right). \end{aligned} \tag{21}$$

Note that the above display is a functional of (X_1, \dots, X_n) , and also a stochastic process indexed by x . In what follows, we show that it is asymptotically a Gaussian process. The distribution of (21) is not straightforward to obtain. The strategy is to first consider a slightly different quantity and then connect it to (21). For each $(x_{(r)}, r)$ such that $|x_{(r)} - \alpha_p| \leq n^{-1/2} \log n$ and $|r/n - G(\alpha_p)| \leq n^{-1/2} \log n$, conditional on $X_{(r)} = x_{(r)}$, $X_{(r+1)}, \dots, X_{(n)}$ are equal in distribution to the order statistics of $n - r$ i.i.d. samples from $Q(X \in \cdot \mid X > x_{(r)})$. Thanks to the fact that $L(x)$ is locally Lipschitz continuous and (E3), we obtain

$$\begin{aligned} & \Phi\left(-\frac{\sum_{i=r+1}^n L(X_{(i)})I(x_{(r)} < X_{(i)} \leq x_{(r)} + xn^{-1/2})}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}} + O(n^{-1/2})\right) \\ &= \Phi\left(-\frac{L(x_{(r)})}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}} \sum_{i=r+1}^n I\left(X_{(i)} \in \left(x_{(r)}, x_{(r)} + \frac{x}{\sqrt{n}}\right]\right) + O(x^2 n^{-1/2})\right). \end{aligned} \tag{22}$$

In the above inequality, we replace $L(X_{(i)})$ by $L(X_{(r)})$. The error term is

$$O(1) \frac{L'(X_{(r)})xn^{-1/2} \sum_{i=r+1}^n I(x_{(r)} < X_{(i)} \leq x_{(r)} + xn^{-1/2})}{\sqrt{n \operatorname{var} \hat{Q} W_{x,n}}} = O(x^2 n^{-1/2}).$$

Note that (22) equals (21) if $\hat{\alpha}_p = X_{(r)} = x_{(r)}$. For the time being, we proceed by conditioning only on $X_{(r)} = x_{(r)}$ and then further derive the conditional distribution of (21) given $\hat{\alpha}_p = X_{(r)} = x_{(r)}$. Owing to Lemma 5, we further simplify the denominator and (22) becomes

$$\Phi\left(-\frac{L(x_{(r)})}{\sqrt{n \operatorname{var} \hat{Q} W_{0,n}}} \sum_{i=r+1}^n I\left(X_{(i)} \in \left(x_{(r)}, x_{(r)} + \frac{x}{\sqrt{n}}\right]\right) + O(x^2 n^{-1/2+\lambda})\right). \tag{23}$$

Let

$$G_{x_{(r)}}(x) = \frac{G(x_{(r)} + x) - G(x_{(r)})}{1 - G(x_{(r)})} = Q(X \leq x_{(r)} + x \mid X > x_{(r)}).$$

Thanks to the result of strong approximation (see [33]–[35]), given $X_{(r)} = x_{(r)}$, there exists a Brownian bridge $\{B(t) : t \in [0, 1]\}$ such that

$$\begin{aligned} & \sum_{i=r+1}^n I\left(X_{(i)} \in \left(x_{(r)}, x_{(r)} + \frac{x}{\sqrt{n}}\right]\right) \\ &= (n - r)G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right) + \sqrt{n - r}B\left(G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right)\right) + O_p(\log(n - r)), \end{aligned} \tag{24}$$

where the $O_p(\log(n-r))$ is uniform in x . Again, we can localize the event by considering a set in which the error term in the above display is $O(\log(n-r))^2$. We substitute this strong approximation into (23) and obtain

$$\begin{aligned} & \Phi\left(-\frac{L(x_{(r)})}{\sqrt{n \operatorname{var}^Q W_{0,n}}}(n-r)G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right) + O_p(x^2n^{-1/2+\lambda}(\log n)^2)\right) \\ & - \varphi\left(-\frac{L(x_{(r)})}{\sqrt{n \operatorname{var}^Q W_{0,n}}}(n-r)G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right) + O_p(x^2n^{-1/4}(\log n)^2)\right) \\ & \times \frac{(n-r)^{1/2}L(x_{(r)})}{\sqrt{n \operatorname{var}^Q W_{0,n}}}B\left(G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right)\right). \end{aligned} \tag{25}$$

In addition, thanks to condition (C4),

$$\frac{L(x_{(r)})}{\sqrt{n \operatorname{var}^Q W_{0,n}}}(n-r)G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right) = \frac{f(x_{(r)})}{\sqrt{\operatorname{var}^Q W_{0,n}}}x + O(x^{\delta_0+3/2}n^{-1/4-\delta_0/2}). \tag{26}$$

Let

$$\xi(x) = \frac{(n-r)^{1/2}L(x_{(r)})}{\sqrt{n \operatorname{var}^Q W_{0,n}}}B\left(G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right)\right), \tag{27}$$

which is a Gaussian process with mean 0 and covariance function

$$\begin{aligned} \operatorname{cov}(\xi(x), \xi(y)) &= \frac{(n-r)L^2(x_{(r)})}{n \operatorname{var}^Q W_{0,n}}G_{x_{(r)}}\left(\frac{x}{\sqrt{n}}\right)\left(1 - G_{x_{(r)}}\left(\frac{y}{\sqrt{n}}\right)\right) \\ &= (1 + O(n^{-1/4+\lambda/2}))\frac{L(x_{(r)})f(x_{(r)})}{\operatorname{var}^Q W_{0,n}}\frac{x}{\sqrt{n}} \end{aligned}$$

for $0 \leq x \leq y \leq n^\lambda$. Insert (26) and (27) into (25) to obtain, given $X_{(r)} = x_{(r)}$,

$$\begin{aligned} & \int_0^{n^\lambda} 2x\Phi\left(-\frac{\sum_{i=r+1}^n L(X_{(i)})I(x_{(r)} < X_{(i)} \leq x_{(r)} + xn^{-1/2})}{\sqrt{n \operatorname{var}^Q \tilde{W}_{x,n}}} + O(n^{-1/2})\right) dx \\ &= \int_0^{n^\lambda} 2x\Phi\left(-\frac{f(x_{(r)})}{\sqrt{\operatorname{var}^Q W_{0,n}}}x + o(x^2n^{-1/4})\right) dx \\ & \quad - \int_0^{n^\lambda} 2x\varphi\left(-\frac{f(x_{(r)})}{\sqrt{\operatorname{var}^Q W_{0,n}}}x + o(x^2n^{-1/4})\right)\xi(x) dx + o(n^{-1/4}), \end{aligned} \tag{28}$$

where $\varphi(x)$ is the standard Gaussian density function. Owing to Lemma 3 and $|x_{(r)} - \alpha_p| \leq n^{-1/2} \log n$, the first term on the right-hand side of (28) is equal to

$$(1 + o(n^{-1/4})) \int_0^\infty 2x\Phi\left(-\frac{f(\alpha_p)}{\sqrt{\operatorname{var}^Q W_{0,n}}}x\right) dx + o_p(n^{-1/4}). \tag{29}$$

The second term on the right-hand side of (28) multiplied by $n^{1/4}$ converges weakly to a

Gaussian distribution with mean 0 and variance

$$\begin{aligned}
 & \sqrt{n} \int_0^{n^\lambda} \int_0^{n^\lambda} 4xy\varphi\left(-\frac{f(x_{(r)})}{\sqrt{\text{var}^Q W_{0,n}}}x + o(x^2n^{-1/4})\right) \\
 & \quad \times \varphi\left(-\frac{f(x_{(r)})}{\sqrt{\text{var}^Q W_{0,n}}}y + o(x^2n^{-1/4})\right) \text{cov}(\xi(x), \xi(y)) \, dx \, dy \\
 & = (1 + o(1)) \int_0^{n^\lambda} \int_0^{n^\lambda} 4xy\varphi\left(-\frac{f(x_{(r)})}{\sqrt{\text{var}^Q W_{0,n}}}x\right)\varphi\left(-\frac{f(x_{(r)})}{\sqrt{\text{var}^Q W_{0,n}}}y\right) \\
 & \quad \times \frac{L(x_{(r)})f(x_{(r)})}{\text{var}^Q W_{0,n}} \min(x, y) \, dx \, dy \\
 & = (1 + o(1)) \frac{L(\alpha_p)(\text{var}^Q W_{0,n})^{3/2}}{f^4(\alpha_p)\sqrt{\pi}}. \tag{30}
 \end{aligned}$$

To obtain the last step in the above display, we need the following calculation:

$$\begin{aligned}
 \text{var}\left(\int_0^\infty 2x\varphi(x)B(x) \, dx\right) & = \int_0^\infty \int_0^\infty 4xy\varphi(x)\varphi(y) \min(x, y) \, dx \, dy \\
 & = \int_0^\infty \int_0^\infty 4r^3 \cos \theta \sin \theta \min(\cos \theta, \sin \theta) \frac{1}{2\pi} e^{-r^2/2} \, dx \, dy \\
 & = 8 \int_0^{\pi/4} \int_0^\infty r^4 \cos \theta \sin^2 \theta \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} e^{-r^2/2} \, dr \, d\theta \\
 & = 8 \frac{1}{3} \frac{1}{2^{3/2}} \frac{1}{\sqrt{2\pi}} \frac{3}{2} \\
 & = \frac{1}{\sqrt{\pi}}.
 \end{aligned}$$

We insert estimates (29) and (30) into (28) and obtain, conditional on $X_{(r)} = x_{(r)}$,

$$\begin{aligned}
 & n^{1/4} \left[\int_0^{n^\lambda} 2x\Phi\left(-\frac{\sum_{i=r+1}^n L(X_{(i)})I(x_{(r)} < X_{(i)} \leq x_{(r)} + xn^{-1/2})}{\sqrt{n \text{var}^Q \tilde{W}_{x,n}}} + O(n^{-1/2})\right) \, dx \right. \\
 & \quad \left. - \int_0^\infty 2x\Phi\left(-\frac{f(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n}}}x\right) \, dx \right] \\
 & \Rightarrow N\left(0, \frac{1}{2}\tau_p^2\right) \tag{31}
 \end{aligned}$$

as $n - r, r \rightarrow \infty$ subject to the constraint that $|r/n - G^{-1}(\alpha_p)| \leq n^{-1/2} \log n$, where τ_p^2 is defined in the statement of the theorem. We could consider the left-hand side of (31) indexed by r and $n - r$. The limit is in the sense that both r and $n - r$ tend to ∞ in the region where $|r/n - G^{-1}(\alpha_p)| \leq n^{-1/2} \log n$.

The limiting distribution of (31) conditional on $\hat{\alpha}_p = X_{(r)} = x_{(r)}$. We now further consider the limiting distribution of the left-hand side of (31) conditional on $\hat{\alpha}_p = X_{(r)} = x_{(r)}$. To simplify the notation, let

$$\begin{aligned}
 V_n & = -n^{1/4} \int_0^{n^\lambda} 2x\varphi\left(-\frac{f(\alpha_p)}{\sqrt{n \text{var}^Q W_{0,n}}}x + o(x^2n^{-1/4})\right)L(\alpha_p) \\
 & \quad \times \frac{\sum_{i=r+1}^n I(x_{(r)} < X_{(i)} \leq x_{(r)} + xn^{-1/2}) - (n - r)G_{x_{(r)}}(xn^{-1/2})}{\sqrt{n \text{var}^Q W_{0,n}}} \, dx.
 \end{aligned}$$

Then,

$$\begin{aligned} & \int_0^{n^\lambda} \Phi\left(-\frac{\sum_{i=r+1}^n L(X_{(i)})I(x_{(r)} < X_{(i)} \leq x_{(r)} + xn^{-1/2})}{\sqrt{n \operatorname{var} \hat{Q} \tilde{W}_{x,n}}} + O(n^{-1/2})\right) dx \\ & - \int_0^\infty 2x \Phi\left(-\frac{f(\alpha_p)}{\sqrt{\operatorname{var} Q} W_{0,n}} x\right) dx \\ & = n^{-1/4} V_n + o(n^{-1/4}). \end{aligned}$$

The weak convergence result in (31) says that, for each compact set A ,

$$Q(V_n \in A \mid X_{(r)} = x_{(r)}) \rightarrow P(Z \in A)$$

as $n - r, r \rightarrow \infty$ subject to the constraint that $|r/n - G^{-1}(\alpha_p)| \leq n^{-1/2} \log n$, where Z is a Gaussian random variable with mean 0 and variance $\tau_p^2/2$. Note that $\hat{\alpha}_p = X_{(r)} = x_{(r)}$ is equivalent to

$$0 \leq H = \sum_{i=1}^r L(X_{(i)})(1 - p) - p \sum_{i=r+1}^n L(X_{(i)}) \leq L(x_{(r)}).$$

Let

$$U_n = \sum_{i=r+1}^n L(X_{(i)})I(x_{(r)} < X_{(i)} \leq x_{(r)} + n^{\lambda-1/2}) - n P(x_{(r)} < X \leq x_{(r)} + n^{\lambda-1/2})$$

and

$$B_n = \{|U_n| \leq n^{\lambda/2+1/4} \log n\}.$$

Note that, given the partial sum U_n , H is independent of the X_i s in the interval $(x_{(r)}, x_{(r)} + n^{\lambda-1/2})$ and is therefore independent of V_n . For each compact set A and $A_n = \{V_n \in A\} \cap B_n$, we have

$$\begin{aligned} & Q(A_n \mid \hat{\alpha}_p = X_{(r)} = x_{(r)}) \\ & = \frac{Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)}, A_n)}{Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)})} Q(A_n \mid X_{(r)} = x_{(r)}) \\ & = E Q \left[\frac{Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)}, U_n)}{Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)})} \mid X_{(r)} = x_{(r)}, A_n \right] \\ & \quad \times Q(A_n \mid X_{(r)} = x_{(r)}). \end{aligned} \tag{32}$$

The second step of the above equation uses the fact that, on the set B_n ,

$$Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)}, U_n) = Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)}, U_n, A_n).$$

Note that U_n depends only on the X_i s in $(x_{(r)}, x_{(r)} + n^{\lambda-1/2})$, while H is the weighted sum of all the samples. Therefore, on the set $B_n = \{|U_n| \leq n^{\lambda/2+1/4} \log n\}$,

$$\frac{Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)}, U_n)}{Q(0 \leq H \leq L(x_{(r)}) \mid X_{(r)} = x_{(r)})} = 1 + o(1), \tag{33}$$

and the $o(1)$ is uniform in B_n . The rigorous proof of the above approximation can be straightforwardly developed using the Edgeworth expansion of density functions, but is tedious, so we

omit it. We substitute (33) into (32). Note that $Q(B_n | X_{(r)} = x_{(r)}) \rightarrow 1$ and we obtain, for each A ,

$$Q(V_n \in A | \hat{\alpha}_p = X_{(r)} = x_{(r)}) - Q(V_n \in A | X_{(r)} = x_{(r)}) \rightarrow 0.$$

Thus, we obtain, conditional on $\hat{\alpha}_p = X_{(r)}$, $|\hat{\alpha}_p - \alpha_p| \leq n^{-1/2} \log n$, $|r/n - G^{-1}(\alpha_p)| \leq n^{-1/2} \log n$, and, as $n \rightarrow \infty$,

$$\begin{aligned} & n^{1/4} \left[\int_0^{n^\lambda} \hat{Q} \left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}} \right) dx - \int_0^\infty 2x \Phi \left(-\frac{f(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n}}} x \right) dx \right] \\ &= n^{1/4} \left[\int_0^{n^\lambda} \Phi \left(-\frac{\sum_{i=r+1}^n L(X_{(i)}) I(\hat{\alpha}_p < X_{(i)} \leq \hat{\alpha}_p + xn^{-1/2})}{\sqrt{n \text{var}^Q \tilde{W}_{x,n}}} + O(n^{-1/2}) \right) dx \right. \\ &\quad \left. - \int_0^\infty 2x \Phi \left(-\frac{f(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n}}} x \right) dx \right] + o_p(1) \\ &= V_n + o_p(1) \\ &\Rightarrow N \left(0, \frac{\tau_p^2}{2} \right). \end{aligned}$$

Together with (E2), this convergence indicates that, asymptotically, the bootstrap variance estimator is independent of $\hat{\alpha}_p$. Therefore, the unconditional asymptotic distribution is given by

$$\begin{aligned} & n^{1/4} \left[\int_0^{n^\lambda} \hat{Q} \left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}} \right) dx - \int_0^\infty 2x \Phi \left(-\frac{f(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n}}} x \right) dx \right] \\ &\Rightarrow N \left(0, \frac{\tau_p^2}{2} \right). \end{aligned} \tag{34}$$

Using exactly the same argument, the asymptotic distribution of the negative part of the integral is given by

$$\begin{aligned} & n^{1/4} \left[\int_0^{n^\lambda} \hat{Q} \left(\hat{\alpha}_p(\mathbf{Y}) < \hat{\alpha}_p - \frac{x}{\sqrt{n}} \right) dx - \int_0^\infty 2x \Phi \left(-\frac{f(\alpha_p)}{\sqrt{\text{var}^Q W_{0,n}}} x \right) dx \right] \\ &\Rightarrow N \left(0, \frac{\tau_p^2}{2} \right). \end{aligned} \tag{35}$$

Using a conditional independence argument, we find that the negative and positive parts of the integral are asymptotically independent. Putting together the results in Theorem 1, (15), (16), (34), (35), and the moment calculations of Gaussian distributions, we conclude that

$$\begin{aligned} \hat{\sigma}_n^2 &= \int_0^\infty 2x [\hat{Q}(\hat{\alpha}_p(\mathbf{Y}) < \hat{\alpha}_p - x) + \hat{Q}(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + x)] dx \\ &= \frac{1}{n} \int_0^\infty 2x \left[\hat{Q} \left(\hat{\alpha}_p(\mathbf{Y}) < \hat{\alpha}_p - \frac{x}{\sqrt{n}} \right) + \hat{Q} \left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}} \right) \right] dx \\ &\stackrel{D}{=} \frac{\text{var}^Q(W_p)}{nf(\alpha_p)^2} + Zn^{-5/4} + o(n^{-5/4}) \\ &= \sigma_n^2 + Zn^{-5/4} + o(n^{-5/4}), \end{aligned}$$

where $Z \sim N(0, \tau_p^2)$.

6. Proof of Theorem 3

Lemma 6. *Under the conditions of Theorem 3, we have*

$$E^Q(LR^\gamma(X); X < \alpha_p) = \frac{c(\theta) + o(1)}{\gamma\sqrt{m}} e^{-m\gamma I}.$$

We clarify that $LR^\gamma(X) = (LR(X))^\gamma$ is the γ th moment of the likelihood ratio. With this result, if we choose $m^3 = O(n)$, it is sufficient to guarantee that, with probability tending to 1, the ratio between the empirical moments and the theoretical moments are within an ε distance from 1. Thus, the localization results (Lemma 4) are in place.

Lemma 7. *Under the conditions of Theorem 3, for each $\gamma > 0$, there exist constants δ (sufficiently small), u_γ , and l_γ such that*

$$l_\gamma \delta \leq \frac{E^Q(LR^\gamma(x); \alpha_p < X \leq \alpha_p + \delta)}{E^Q(LR^\gamma(x); X < \alpha_p)} \leq u_\gamma \delta$$

for all sufficiently large m .

The proofs of the above two lemmas are standard and use the exponential change of measure and Edgeworth’s expansion. We omit the details.

Proof of Theorem 3. The proof of this theorem is very similar to those of Theorems 1 and 2. The only difference is that we need to keep in mind that there is another parameter m that tends to ∞ . Therefore, the main task of this proof is to provide a careful analysis and establish a sufficiently large n so that similar asymptotic results hold as m tends to ∞ in a slower manner than n .

From a technical point of view, the main reason why we need to choose $m^3 = O(n)$ is that we use the Berry–Esseen bound in the region $[0, n^\lambda]$ to approximate the distribution of $\sqrt{n}(\hat{\alpha}_p - \alpha_p)$. In order to have the approximation hold (see case 2 of part 1 below), it is necessary to have $m^{1/4} = o(n^\lambda)$. On the other hand, the error term of the Berry–Esseen bound requires that n^λ cannot be too large. The order $m^3 = O(n)$ is sufficient to guarantee both.

The proof consists of two parts. In part 1 we establish similar results as those in Theorem 1; in part 2 we establish the corresponding results given in Theorem 2.

Part 1. We now proceed to establish the asymptotic mean and variance of the weighted quantile estimator. Recall that in the proof of Theorem 1 we developed the approximations of the tail probabilities of the quantile estimator and used Lemma 1 to conclude the proof. In particular, we approximated the right tail of $\hat{\alpha}_p$ in three different regions. We go through the three cases carefully for some $\max(1/4(\alpha - 2), \frac{1}{10}) < \lambda < \frac{1}{8}$ (recall that $L(X)$ has at least α th moment under Q).

Case 1: $0 < x \leq n^\lambda$. We approximate the tail probability using the Berry–Esseen bound:

$$Q(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) = \Phi\left(\frac{-(F(\alpha_p + xn^{-1/2}) - p)}{\sqrt{\text{var}^Q W_{x,n,1}/n}}\right) + D_1(x). \tag{36}$$

Here

$$W_{x,n,i} = LR(X_i)I(X_i \leq \alpha_p + xn^{-1/2}) - F(\alpha_p + xn^{-1/2})$$

and

$$|D_1(x)| \leq \frac{c E |W_{x,n,1}|^3}{(\text{var}^Q W_{x,n,1})^{3/2}} n^{-1/2} = O(1) \frac{m^{1/4}}{n^{1/2}}.$$

The equality in the approximation of $D_1(x)$ follows from Lemma 6. In the current case, $W_{x,n,i}$ has all the moments, that is, α can be chosen arbitrarily large.

Case 2: $n^\lambda < x \leq c\sqrt{n}$. Applying Theorem 2.18 of [16], for each $\delta > 0$, there exist $\kappa_1(\delta)$ and $\kappa_2(\delta)$ such that

$$\begin{aligned} \mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) &= \mathbb{Q}\left(\sum_{i=1}^n W_{x,n,i} < n\left(p - F\left(\alpha_p + \frac{x}{\sqrt{n}}\right)\right)\right) \\ &\leq \left(\frac{3}{1 + \delta n \operatorname{var}(W_{x,n,1})^{-1}(p - F(\alpha_p + xn^{-1/2}))^2}\right)^{1/\delta} \\ &\quad + \mathbb{Q}\left(\min_{i=1,\dots,n} W_{x,n,i} < \delta n(p - F(\alpha_p + xn^{-1/2}))\right) \\ &\leq \kappa_1(\delta)(m^{-1/4}x)^{-2/\delta} + \frac{\kappa_2(\delta)n^{-\alpha/2+1}}{\sqrt{m}}(m^{-1/2}x)^{-\alpha}. \end{aligned}$$

The last inequality follows from Lemma 7 with $\gamma = 1$. Given that we can choose δ arbitrarily small and α arbitrarily large (thanks to the steepness), and $m^3 = O(n)$, we obtain

$$\mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) = O(x^{-\alpha+\varepsilon}),$$

where ε can be chosen arbitrarily small. Thus,

$$\int_{n^\lambda}^{c\sqrt{n}} x \mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) dx = o(n^{-1/4}).$$

Case 3: $x > c\sqrt{n}$. Similarly to the proof of Theorem 1, for $c\sqrt{n} < x \leq n^{\alpha/6-\varepsilon/6}$,

$$\mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) \leq \kappa_3 x^{-3}.$$

For $x^{\beta-3} > n^{1+2\beta/3}$ (recall that X has at least β th moment under \mathbb{Q}),

$$\mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) \leq \mathbb{Q}\left(\sup_i X_i > \alpha_p + \frac{x}{\sqrt{n}}\right) \leq O(1)n^{1+\beta/2}m^{\beta/2}x^{-\beta} = O(x^{-3}).$$

Given the steepness of $\varphi(\theta)$, α and β can be chosen arbitrarily large. We then have $1 = \alpha/6 \geq (1 + 2\beta/3)/(\beta - 3)$. Thus, we conclude that

$$\mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) = O(x^{-3})$$

for all $x > c\sqrt{n}$ and

$$\int_{c\sqrt{n}}^\infty x \mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) dx = O(n^{-1/2}).$$

Summarizing the three cases, we have

$$\begin{aligned} &\int_0^\infty x \mathbb{Q}(\sqrt{n}(\hat{\alpha}_p - \alpha_p) > x) dx \\ &= \int_0^{n^\lambda} x \Phi\left(\frac{-(F(\alpha_p + xn^{-1/2}) - p)}{\sqrt{\operatorname{var}^{\mathbb{Q}} W_{x,n,1}/n}}\right) dx + o(m^{1/4}n^{-1/4}) \\ &= \int_0^{n^\lambda} x \Phi\left(-\frac{f(\alpha_p)x}{\sqrt{\operatorname{var}^{\mathbb{Q}} W_{0,n,1}}}\right) dx + o(m^{1/4}n^{-1/4}). \end{aligned}$$

For the last step, we used the fact that $\text{var}^Q W_{x,n,1} = (1 + O(xn^{-1/2})) \text{var}^Q W_{0,n,1}$, which is an application of Lemma 7. Using Lemma 1, we obtain

$$E^Q(\hat{\alpha}_p) = \alpha_p + o(m^{1/4}n^{-3/4}), \quad \sigma_n^2 = \text{var}^Q(\hat{\alpha}_p) = \frac{\text{var}^Q(W_p)}{nf(\alpha_p)^2} + o(m^{1/4}n^{-5/4}),$$

where

$$W_p = LR(X)I(X \leq \alpha_p)$$

and the convergence is uniform in m when $m^3 = O(n)$. Note that we aim to show that the bootstrap variance estimator converges to σ_n^2 at a rate of $O(m^{5/8}n^{-5/4})$. Then, we can basically treat the above approximations as true values in the derivations for the bootstrap variance estimator.

Part 2. We now proceed to the discussion of the distribution of the bootstrap variance estimator. The proof needs to go through three similar cases as in part 1 where we derived the approximation of the mean and variance of $\hat{\alpha}_p$ under Q . The difference is that we need to handle the empirical measure \hat{Q} instead of Q . As we explained after the statement of Lemma 6, the localization conditions (Lemma 4) are satisfied when $m^3 = O(n)$. Let the set C_n be as defined in Lemma 4. The analyses of these three cases are identical to those in part 1. We obtain similar results as in part 1, i.e.

$$\int_{n^\lambda}^\infty x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) dx = o(n^{-1/4}). \tag{37}$$

We omit the detailed analysis. In what follows, we focus on revisiting case 1 in the proof of Theorem 2, which is the leading term of the asymptotic result. Then, we continue the derivation from (17). On the set C_n and with a similar result (for the empirical measure \hat{Q}) as in (36), we have

$$\begin{aligned} & \int_0^{n^\lambda} x \hat{Q}\left(\hat{\alpha}_p(\mathbf{Y}) > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) dx \\ &= \int_0^{n^\lambda} x \Phi\left(-\frac{\sum_{i=1}^n LR(X_i)I(X_i \leq \hat{\alpha}_p + xn^{-1/2}) - np}{\sqrt{n \text{var}^{\hat{Q}} W_{x,n}}}\right) dx + o(m^{1/4}n^{-1/4}). \end{aligned} \tag{38}$$

We take a closer look at the above Gaussian probability. Using the same arguments as in (18) and (19), we obtain

$$\begin{aligned} & \Phi\left(-\frac{\sum_{i=1}^n LR(X_i)I(X_i \leq \hat{\alpha}_p + xn^{-1/2}) - np}{\sqrt{n \text{var}^{\hat{Q}} W_{x,n}}}\right) \\ &= \Phi\left(-\frac{\sum_{i=1}^n LR(X_i)I(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + xn^{-1/2})}{\sqrt{n \text{var}^{\hat{Q}} W_{x,n}}} + O(m^{1/4}n^{-1/2})\right). \end{aligned}$$

We replace $LR(X_i)$ by $LR(\hat{\alpha}_p)$ and obtain

$$\Phi\left(\frac{-LR(\hat{\alpha}_p) \sum_{i=1}^n I(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + xn^{-1/2})}{\sqrt{n \text{var}^{\hat{Q}} W_{x,n}}} + O(x^2m^{-1/4}n^{-1/2} + m^{1/4}n^{-1/2})\right).$$

Lastly, we replace the empirical variance in the denominator by the theoretical variance. Similarly to the development in Lemma 5, we obtain the estimates

$$|\text{var}^Q W_{0,n} - \text{var}^{\hat{Q}} W_{x,n}| = (1 + O_p(m^{1/4}n^{-1/2} + xn^{-1/2})) \text{var}^Q W_{0,n}$$

for $|x| < n^\lambda$. Since we are deriving weak convergence results, we can always localize the events so that $O_p(\cdot)$ can be replaced by $O(\cdot)$. Then, the Gaussian probability is approximated by

$$\Phi\left(\frac{-LR(\hat{\alpha}_p) \sum_{i=1}^n I(\hat{\alpha}_p < X_i \leq \hat{\alpha}_p + xn^{-1/2})}{\sqrt{n \text{var}^Q W_{0,n}}} + \zeta(x, m, n)\right),$$

where $\zeta(x, m, n) = O(x^2m^{-1/4}n^{-1/2} + xn^{-1/2} + m^{1/4}n^{-1/2})$. Using the strong approximation of empirical processes in (24) and the Lipschitz continuity of the density function, we can further approximate the above probability by

$$\begin{aligned} &\Phi\left(\frac{f(\hat{\alpha}_p)}{\sqrt{\text{var}^Q W_{0,n}}}x + \zeta(x, m, n)\right) \\ &- \varphi\left(\frac{f(\hat{\alpha}_p)}{\sqrt{\text{var}^Q W_{0,n}}}x + \zeta(x, m, n)\right) \frac{(n-r)^{1/2}L(\hat{\alpha}_p)}{\sqrt{n \text{var}^Q W_p}} B\left(G_{\hat{\alpha}_p}\left(\frac{x}{\sqrt{n}}\right)\right), \end{aligned}$$

where $B(t)$ is a standard Brownian bridge and

$$G_y(x) = Q(X \leq y + x | X > y).$$

Note that

$$\begin{aligned} &\int_0^{n^\lambda} 2x \Phi\left(\frac{f(\hat{\alpha}_p)}{\sqrt{\text{var}^Q W_{0,n}}}x + \zeta(x, m, n)\right) dx \\ &= \int_0^{n^\lambda} 2x \varphi\left(\frac{f(\hat{\alpha}_p)}{\sqrt{\text{var}^Q W_{0,n}}}x\right) dx + O(m^{3/4}n^{-1/2}). \end{aligned}$$

We write

$$Z_n = -n^{1/4} \int_0^{n^\lambda} 2x \varphi\left(\frac{f(\hat{\alpha}_p)}{\sqrt{\text{var}^Q W_{0,n}}}x + \zeta(x, m, n)\right) \frac{(n-r)^{1/2}L(\hat{\alpha}_p)}{\sqrt{n \text{var}^Q W_p}} B\left(G_{\hat{\alpha}_p}\left(\frac{x}{\sqrt{n}}\right)\right) dx.$$

The calculation of the asymptotic distribution of Z_n is completely analogous to (30), so we omit it. Putting all these results together, the integral in (38) can be written as

$$\begin{aligned} &\int_0^{n^\lambda} 2x \hat{Q}\left(\hat{\alpha}_p(Y) > \hat{\alpha}_p + \frac{x}{\sqrt{n}}\right) dx \\ &= \int_0^\infty 2x \Phi\left(\frac{f(\hat{\alpha}_p)}{\sqrt{\text{var}^Q W_{0,n}}}x\right) dx + \frac{Z_n}{n^{1/4}} + o(m^{5/8}n^{-1/4}), \end{aligned}$$

where

$$\frac{Z_n}{\sqrt{\tau_p/2}} \Rightarrow N(0, 1), \quad \tau_p^2 = 2\pi^{-1/2}L(\alpha_p)f(\alpha_p)^{-4}(\text{var}^Q(W_p))^{3/2} = O(m^{5/4})$$

as $n \rightarrow \infty$ uniformly when $m^3 = O(n)$. The derivation for $\int_0^{n^\lambda} x \hat{Q}(\hat{\alpha}_p(Y) < \hat{\alpha}_p - x/\sqrt{n}) dx$ is analogous to the positive part. Together with (37), we conclude the proof.

References

- [1] ADLER, R. J., BLANCHET, J. H. AND LIU, J. C. (2009). Efficient Monte Carlo for large excursions of Gaussian random fields. Preprint.
- [2] ASMUSSEN, S. AND GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.
- [3] ASMUSSEN, S. AND KROESE, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. Appl. Prob.* **38**, 545–558.
- [4] ASMUSSEN, S., BINSWANGER, K. AND HØJGAARD, B. (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli* **6**, 303–322.
- [5] BABU, G. J. (1986). A note on bootstrapping the variance of sample quantile. *Ann. Inst. Statist. Math.* **38**, 439–443.
- [6] BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580.
- [7] BLANCHET, J. H. (2009). Efficient importance sampling for binary contingency tables. *Ann. Appl. Prob.* **19**, 949–982.
- [8] BLANCHET, J. AND GLYNN, P. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Prob.* **18**, 1351–1378.
- [9] BLANCHET, J. H. AND LIU, J. (2008). State-dependent importance sampling for regularly varying random walks. *Adv. Appl. Prob.* **40**, 1104–1128.
- [10] BLANCHET, J. AND LIU, J. (2010). Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *J. Appl. Prob.* **47**, 301–322.
- [11] BLANCHET, J., GLYNN, P. AND LIU, J. C. (2007). Efficient rare event simulation for multiserver queues. Preprint.
- [12] BLANCHET, J., GLYNN, P. AND LIU, J. C. (2007). Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. *Queueing Systems* **57**, 99–113.
- [13] BUCKLEW, J. A. (2004). *Introduction to Rare Event Simulation*. Springer, New York.
- [14] CHU, F. AND NAKAYAMA, M. K. (2010). Confidence intervals for quantiles and value-at-risk when applying importance sampling. In *Proc. of the 2010 Winter Simulation Conference*, IEEE, pp. 2751–2761.
- [15] DAVID, H. A. AND NAGARAJA, H. N. (2003). *Order Statistics*, 3rd edn. Wiley-Interscience, Hoboken, NJ.
- [16] DE LA PEÑA, V. H., LAI, T. L. AND SHAO, Q.-M. (2009). *Self-Normalized Processes*. Springer, Berlin.
- [17] DUPUIS, P. AND WANG, H. (2005). Dynamic importance sampling for uniformly recurrent Markov chains. *Ann. Appl. Prob.* **15**, 1–38.
- [18] DUPUIS, P., LEDER, K. AND WANG, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.* **17**, 21 pp.
- [19] DUPUIS, P., SEZER, A. L. AND WANG, H. (2007). Dynamic importance sampling for queueing networks. *Ann. Appl. Prob.* **17**, 1306–1346.
- [20] DURRETT, R. (2010). *Probability: Theory and Examples*, 4th edn. Cambridge University Press.
- [21] FALK, M. (1986). On the estimation of the quantile density function. *Statist. Prob. Lett.* **4**, 69–73. (Correction: **4** (1986), 217.)
- [22] GHOSH, M., PARR, W. C., SINGH, K. AND BABU, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12**, 1130–1135.
- [23] GLASSERMAN, P. AND LI, J. (2005). Importance sampling for portfolio credit risk. *Manag. Sci.* **51**, 1643–1656.
- [24] GLASSERMAN, P., HEIDELGERGER, P. AND SHAHABUDDIN, P. (1999). Importance sampling and stratification for value-at-risk. In *Computational Finance 1999* (Proc. 6th Internat. Conf. Comput. Finance), MIT Press, Cambridge, MA, pp. 7–24.
- [25] GLASSERMAN, P., HEIDELBERGER, P. AND SHAHABUDDIN, P. (2000). Variance reduction techniques for estimating value-at-risk. *Manag. Sci.* **46**, 1349–1364.
- [26] GLASSERMAN, P., HEIDELBERGER, P. AND SHAHABUDDIN, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance* **12**, 239–269.
- [27] GLYNN, P. W. (1996). Importance sampling for Monte Carlo estimation of quantiles. In *Mathematical Methods in Stochastic Simulation and Experimental Design: Proc. 2nd St. Petersburg Workshop on Simulation*, Publishing House of Saint Petersburg, pp. 180–185.
- [28] HALL, P. AND MARTIN, M. A. (1988). Exact convergence rate of bootstrap quantile variance estimator. *Prob. Theory Relat. Fields* **80**, 261–268.
- [29] HULT, H. AND SVENSSON, J. (2009). Efficient calculation of risk measures by importance sampling – the heavy tailed case. Preprint.
- [30] HULT, H. AND SVENSSON, J. (2009). Efficient calculation of risk measures by importance sampling – the heavy tailed case. Preprint.
- [31] JUNEJA, S. AND SHAHABUDDIN, P. (2002). Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Trans. Model. Comput. Simul.* **12**, 94–118.
- [32] KISH, L. (1965). *Survey Sampling*. John Wiley, New York.
- [33] KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitsth.* **32**, 111–131.

- [34] KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1975). Weak convergence and embedding. In *Limit Theorems of Probability Theory* (Colloq. Math. Soc. János Bolyai **11**), North-Holland, Amsterdam, pp. 149–165.
- [35] KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitsth.* **34**, 33–58.
- [36] LOHR, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA.
- [37] MARITZ, J. S. AND JARRETT, R. G. (1978). A note on estimating the variance of the sample median. *J. Amer. Statist. Assoc.* **73**, 194–196.
- [38] MCKEAN, J. W. AND SCHRADER, R. M. (1984). A comparison of methods for studentizing the sample median. *Commun. Statist. Simul. Comput.* **13**, 751–773.
- [39] SADOWSKY, J. S. (1996). On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Prob.* **6**, 399–422.
- [40] SHEATHER, S. J. (1986). A finite sample estimate of the variance of the sample median. *Statist. Prob. Lett.* **4**, 337–342.
- [41] SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4**, 673–684.
- [42] SUN, L. AND HONG, L. J. (2009). A general framework of importance sampling for value-at-risk and conditional value-at-risk. In *Proc. 2009 Winter Simulation Conference*, pp. 415–422.
- [43] SUN, L. AND HONG, L. J. (2010). Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operat. Res. Lett.* **38**, 246–251.
- [44] WANG, R.-H., LIN, S.-K. AND FUH, C.-D. (2009). An importance sampling method to evaluate value-at-risk for asset with jump risk. *Asia-Pacific J. Financial Studies* **38**, 745–772.