Order Number 9518379

Bootstrapping extremes of random variables

Fukuchi, Jun-ichiro, Ph.D.

Iowa State University, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Bootstrapping extremes of random variables

by

Jun-ichiro Fukuchi

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1994

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

1.1 Bootstrap

Efron (1979) introduced the bootstrap method as a means of estimating the sampling distribution of statistics. The bootstrap method is described as follows by Efron (1979). Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with the cdf F and let $\mathbf{X}_n = (X_1, X_2, \cdots, X_n)$. Given a specified random variable $R(\mathbf{X}_n, F)$, possibly depending on both $\mathbf{X}_n$ and unknown distribution F, we wish to estimate the sampling distribution of R based on $\mathbf{X}_n$ itself. The bootstrap method is:

1. Construct the empirical distribution function $F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$ of $\mathbf{X}_n$.

2. With F_n fixed, draw a random sample of size $m = m_n$ with replacement from F_n, say

$$X_i^* \overset{ind}{\sim} F_n, \quad i = 1, 2, \cdots, m.$$

Call this the *bootstrap sample*, and let $\mathbf{X}_n^* = (X_1^*, X_2^*, \cdots, X_m^*)$.

3. Approximate the sampling distribution of $R(\mathbf{X}_n, F)$ by the conditional distribution of $R(\mathbf{X}_n^*, F_n)$ given $\mathbf{X}_n$ (called the bootstrap distribution of $R(\mathbf{X}_n, F)$). n and m are called the sample size and the resample size respectively.

Remark 1.1 Efron (1979) considered only the case $m = n$. Bickel and Freedman (1981) were among the earliest to consider the possibility of m different from n.

The bootstrap method has been having a great impact on both theory and applications of statistics. Many papers have been devoted to the theoretical justifications of the bootstrap and its variants. One of the desired properties of bootstrap is consistency defined as follows:

Definition 1.1 Let ρ be a metric on the space of probabilty measures, which is equivalent to convergence in distribution (e.g., the Prohorov metric). The bootstrap is said to be weakly (strongly) consistent if

$$\rho(P\{R(\mathbf{X}_n, F) \le \cdot\}, P\{R(\mathbf{X}_n^*, F_n) \le \cdot \mid \mathbf{X}_n\}) \to 0 \tag{1.1}$$

in probability (w.p.1).

Note that if the limit cdf of $P\{R(\mathbf{X}_n, F) \le \cdot\}$ exists and is continuous, (1.1) is equivalent to

$$\sup_{x \in \mathbf{R}} \mid P\{R(\mathbf{X}_n, F) \le x\} - P\{R(\mathbf{X}_n^*, F_n) \le x \mid \mathbf{X}_n\} \mid \to 0 \tag{1.2}$$

in probability (w.p.1). Bickel and Freedman (1981) proved the consistency of the bootstrap for the sample mean, for von Mises functionals, for the empirical process, and for the quantile process.

1.2 m out of n Bootstrap

Bickel and Freedman (1981) also showed examples where the bootstrap is inconsistent when $m_n = n$. Those are some U-statistics and the maximum order statistics of random variables. Beran (1982) showed that the bootstrap is inconsistent for

Hodges's super-efficient estimator for the means. Athreya (1987) established the inconsistency of the bootstrap for the sample mean of heavy tailed random variables. He showed that the bootstrap distribution converges in distribution to a random cdf in this case.

Inconsistency of the bootstrap is often fixed by making resample size m smaller than n. Bretagnolle (1983) was the first to find this kind of phenomena. He showed that if m is small compared to n, then the bootstrap for the U-statistic is consistent. For the case of the sample mean of heavy tailed random variables, the bootstrap is weakly consistent if $m = o(n)$ (Athreya, 1986), and it is strongly consistent if $m = o(n(\log\log n)^{-1})$ (Arcones and Gine, 1989). Bickel (1994) called the bootstrap with small resample size the "m out of n bootstrap". Recently similar results have been found by several authors in different problems and the m out of n bootstrap has been applied to associated inference problems. See Datta (1992) for the estimation in the unstable AR(1) process, Athreya, Lahiri and Wu (1993) for a confidence interval of the mean of the heavy tailed distribution, Davis and Wu (1994) for M-estimators in an infinite variance autoregression, and Datta and McCormick (1994) for an extreme value estimator of the autoregressive parameter of AR(1) process.

A closely related method is the subsampling method which is equivalent to the m out of n without replacement bootstrap. Politis and Romano (1992) proposed this method and proved that it can approximate the distribution of the statistic of the form $\tau_n(T_n(\mathbf{X}_n) - \theta)$ which has a non-degenerate weak limit, if the subsampling size b satisfies $b/n \to 0$ and $\tau_b/\tau_n \to 0$. They also proved, as a corollary, that the m out of n with replacement bootstrap is weakly consistent for the statistics of the same form if $\tau_m/\tau_n \to 0$ and $m^2/n \to 0$. These results reveal the wide applicability of the subsampling method and m out of n bootstrap.

1.3 Focus of this thesis

Although the bootstrap is inconsistent for the maximum order statistics, the works of Bretagnolle (1983), Athreya (1986) and Arcones and Gine (1989) mentioned earlier suggest that if the resample size is $m = o(n)$, then the bootstrap could be made consistent for the maximum order statistics. This is indeed the case and it is the focus of this thesis.

As stated precisely in Section 2.1, possible limit distributions of the maximum order statistics of i.i.d. random variables are completely known (called extreme value distributions), but they depend on the population parameter which in practice is unknown. Therefore, the bootstrap method is a good alternative to the extreme value distributions for approximations to the cdf's of normalized maximum.

The organization of this thesis is as follows:

In Chapter 2, after the classical extreme value theory is reviewed, asymptotic properties of the bootstrap for the extremes of i.i.d. random variables are investigated. Inconsistency and consistency of the bootstrap is proved for various resample sizes. Also, results on a smoothed bootstrap are given.

In Chapter 3, some results in Chapter 2 are extended to stationary processes. The moving block bootstrap introduced by Künsch (1989) and Liu and Singh (1992) is defined for the normalized maximum of a stationary process and its consistency is proved.

In Chapter 4, results from Chapter 2 are applied to the problem of constructing confidence intervals of endpoints of a cdf. Some simulation results are given.

2. BOOTSTRAPPING EXTREMES OF I.I.D. RANDOM VARIABLES

2.1 Asymptotic Theory of Extremes of I.I.D. Random Variables

This section presents a review of the asymptotic theory of extremes of i.i.d. random variables. Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with a cdf F and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the corresponding order statistics. Let $F^{-1}(u) := \inf\{x : F(x) \geq u\}$ be the left continuous inverse of F and C_G be the set of continuity points of a function G.

The following theorem is the most basic in the extreme value theory.

Theorem 2.1 (Gnedenko, 1943) *Suppose that there exists $a_n > 0$, $b_n \in \mathbf{R}$, $n \geq 1$ such that*

$$P\{a_n^{-1}(X_{n:n} - b_n) \leq x\} = F^n(a_n x + b_n) \to G(x) \tag{2.1}$$

for each $x \in C_G$ as $n \to \infty$, where G is nondegenerate cdf . Then G is of the type of one of the following three classes.

$$(i) \qquad \Lambda(x) = \exp(-e^{-x}) \quad x \in \mathbf{R},$$

$$(ii) \qquad \Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \end{cases}$$

for some $\alpha > 0$,

$$(iii) \qquad \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & x \le 0 \\ 1 & x > 0, \end{cases}$$

for some $\alpha > 0$.

We say that F belongs to the domain of attraction of G (written $F \in D(G)$) if (2.1) holds for some $a_n, b_n \in R$.

The following theorem gives necessary and sufficient conditions for the cdf F to belong to the domain of attraction of each of the three types. Let $\theta_F = \sup\{x : F(x) < 1\}$ be the upper endpoint of F. Then

Theorem 2.2 (Gnedenko, 1943) *(i)* $F \in D(\Lambda)$ *iff there exists strictly positive function* g *such that*

$$lim_{t \uparrow \theta_F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}.$$

(ii) $F \in D(\Phi_\alpha)$ *iff* $\theta_F = \infty$ *and* $\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}$. *for each* $x > 0$

(iii) $F \in D(\Psi_\alpha)$ *iff* $\theta_F < \infty$ *and* $\lim_{h \downarrow 0} \frac{1 - F(\theta_F - xh)}{1 - F(\theta_F - h)} = x^{\alpha}$. *for each* $x > 0$

The normal distribution belongs to $D(\Lambda)$, the Cauchy distribution belongs to $D(\Phi_\alpha)$ and the uniform distribution belong to $D(\Psi_\alpha)$.

Theorem 2.3 (Gnedenko, 1943; de Haan, 1970) *Normalizing constants* a_n, b_n *in (2.1) may be chosen as:*

$$\begin{aligned} &(i) \quad \text{if } F \in D(\Lambda), \quad a_n = F^{-1}(1 - \frac{1}{en}) - \gamma_n, \quad b_n = \gamma_n, \\ &(ii) \quad \text{if } F \in D(\Phi_\alpha), \quad a_n = \gamma_n, \quad b_n = 0, \\ &(iii) \quad \text{if } F \in D(\Psi_\alpha), \quad a_n = \theta_F - \gamma_n, \quad b_n = \theta_F, \end{aligned} \tag{2.2}$$

where $\gamma_n = F^{-1}(1 - \frac{1}{n})$

2.2 Bootstrapping Extremes when m=n

This section presents results on asymptotic behaviors of the bootstrap distribution when resample size m is of the same order of sample size n. Bickel and Freedman (1981) showed that the bootstrap of the maximum of i.i.d. uniform random variables is inconsistent. Angus (1993) further showed that the bootstrap distribution converges in distribution to a random distribution and obtained the explicit form of the limit when normalizing constants are estimated. In this section, it is shown that, when normalizing constants are known, the bootstrap distribution converges in distribution to a random distribution.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables on it with a cdf $F \in D(G)$ where $G = \Lambda, \Phi_\alpha$ or Ψ_α. Let $m = m(n) \in \mathbf{N}$ be such that $m \to \infty$ as $n \to \infty$. Given $\mathbf{X}_n \equiv (X_1, X_2, \cdots, X_n)$, let $Y_1, Y_2, \cdots Y_m$ be conditionally i.i.d. random variables with the distribution

$$P(Y_1 = X_j | \mathbf{X}_n) = \frac{1}{n}, \quad j = 1, 2, \cdots, n,$$

i.e. $(Y_1, Y_2, \cdots, Y_m)$ is a random sample of size m from the empirical distribution $F_n(\cdot) = n^{-1} \sum_{i=1}^{n} I(X_i \leq \cdot)$ of $\mathbf{X}_n$. Let $Y_{1:m} \leq Y_{2:m} \leq \cdots \leq Y_{m:m}$ be the corresponding order statistics. Now, define

$$G_n(x) = P\{a_n^{-1}(X_{n:n} - b_n) \leq x\}, \tag{2.3}$$

$$H_{n,m}(x, \omega) = P\{a_m^{-1}(Y_{m:m} - b_m) \leq x \mid \mathbf{X}_n\}. \tag{2.4}$$

$H_{n,m}(x,\omega)$ is called the *bootstrap distribution* of $a_n^{-1}(X_{1:n} - b_n)$. n and m are called the *sample size* and *resample size* respectively.

Note that (2.1) is equivalent to

$$n\{1 - F(a_n x + b_n)\} \to c(x) \equiv -\log G(x), \tag{2.5}$$

for each $x \in C_G$. Note also that

$$\begin{aligned} H_{n,m}(x,\omega) &= F_n^m(a_m x + b_m) \\ &= \{1 - m^{-1}m(1 - F_n(a_m x + b_m))\}^m. \end{aligned} \tag{2.6}$$

Now define for each $A \in \mathcal{B}(\mathbf{R})$, the Borel σ-algebra on $\mathbf{R}$

$$T_{n,m}(A,\omega) = \#\{j : 1 \le j \le n, a_m^{-1}(X_j - b_m) \in A\}, \tag{2.7}$$

where $\#$ denotes the cardinality of a set. It suffices to investigate asymptotic behavior of $T_{n,m}(\cdot,\omega)$ since $T_{n,m}((x,\infty),\omega) = n\{1 - F_n(a_m x + b_m)\}$. We need some preliminaries before stating asymptotic properties of $T_{n,m}(\cdot,\omega)$. We follow Chapter 3 of Resnick (1987) to describe point processes and their weak convergence.

Definition 2.1 A measure ν on $\mathcal{B}(\mathbf{R})$ is said to be *Radon* if $\nu(K) < \infty$ for each compact set $K \in \mathbf{R}$.

For $x \in \mathbf{R}$, define the measure ϵ_x on $\mathcal{B}(\mathbf{R})$ by

$$\epsilon_x = \begin{cases} 1 & x \in A \\ 0 & x \in A^C. \end{cases}$$

Definition 2.2 A measure ν on $\mathcal{B}(\mathbf{R})$ is called a *point measure* if it is Radon and

$$\nu = \sum_{i=1}^{\infty} \epsilon_{x_i}$$

for some countable collection $\{x_i, i \geq 1\}$ of points in $\mathbf{R}$.

Definition 2.3 Let $M_p(\mathbf{R})$ be the set of all point measures defined on $\mathbf{R}$, and $\mathcal{M}_p(\mathbf{R})$ be the smallest $\sigma-$algebra of subsets of $M_p(\mathbf{R})$ making all evaluation map $m \to m(A)$ measurable for each $A \in \mathcal{B}(\mathbf{R})$, then a measurable map $N : \Omega \to M_p(\mathbf{R})$ is called a *point process* on $\mathbf{R}$.

According to Definition 2.3, $T_{n,m}$ defined in (2.7) is a point process on $\mathbf{R}$.

Definition 2.4 Let μ be a Radon measure on $\mathcal{B}(\mathbf{R})$. A point process N on $\mathbf{R}$ is called a *Poisson random measure with mean measure* μ if N satisfies:

(i) For any $A \in \mathcal{B}(\mathbf{R})$ and any integer $k \geq 0$,

$$P\{N(A) = k\} = \begin{cases} \exp\{-\mu(A)\}\frac{\mu(A)^k}{k!} & \text{if } \mu(A) < \infty \\ 0 & \text{if } \mu(A) = \infty, \end{cases}$$

(ii) For any $k \geq 1$, if $A_1, \cdots, A_k$ are disjoint sets in $\mathbf{R}$, then $N(A_i), i = 1, \cdots, k$, are independent.

Definition 2.5 Let $C_K^+(\mathbf{R})$ be the set of all continuous, nonnegative real valued functions on $\mathbf{R}$ with compact support. For $\mu_n, \mu \in M_p(\mathbf{R})$ we say μ_n *converges vaguely to* μ (written $\mu_n \xrightarrow{v} \mu$) if

$$\int_{\mathbf{R}} f(x)\,\mu_n(dx) \to \int_{\mathbf{R}} f(x)\,\mu(dx)$$

for each $f \in C_K^+(\mathbf{R})$.

A topology on $M_p(\mathbf{R})$ giving this notion of convergence is called the *vague topology* on $M_p(\mathbf{R})$ and is known to be metrizable as a complete, separable metric space (cf. Resnick, 1987, pp. 147).

Definition 2.6 Let $\{N_n, n \geq 0\}$ be a sequence of random elements in $M_p(\mathbf{R}^l)$ (point processes on $\mathbf{R}$). We say that N_n *converges to* N_0 *weakly* (written $N_n \Rightarrow N_0$) if

$$E(f(N_n)) \to E(f(N_0)),$$

for each $f \in C_B(M_p(\mathbf{R}^l))$ where $C_B(M_p(\mathbf{R}^l))$ is the set of all bounded, continuous real valued functions on $M_p(\mathbf{R}^l)$.

The next theorem shows that if m=n, $H_{n,m}(x,\omega)$ has a random limit and thus the naive bootstrap distribution fails to approximate $G_n(x)$. Let PRM(μ) denote a Poisson random measure with mean measure μ.

Theorem 2.4 *Suppose that (2.1) holds.*

(i) If $G = \Lambda$, let ν be a measure on $\mathcal{B}((-\infty,\infty])$ determined by $\nu(x,\infty] = e^{-x}$, $x \in \mathbf{R}$, and T be a PRM(ν) on $[\infty,\infty)$, and let $H(x) = \exp\{-T((x,\infty))\}$, then for any $x_i \in \mathbf{R}$, $i = 1,\cdots,r$,

$$(H_{n,n}(x_i,\cdot), i = 1,2,\cdots,r) \xrightarrow{d} (H(x_i,\cdot), i = 1,2,\cdots,r). \tag{2.8}$$

(ii) If $G = \Phi_\alpha$, let ν be a measure on $\mathcal{B}((0,\infty])$ determined by $\nu(x,\infty] = x^{-\alpha}$, $x > 0$, and T be a PRM(ν) on $(0,\infty]$, and let $H(x) = \exp\{-T((x,\infty))\}$, then for any $x_i > 0$, $i = 1,\cdots,r$,

$$(H_{n,n}(x_i,\cdot), i = 1,2,\cdots,r) \xrightarrow{d} (H(x_i,\cdot), i = 1,2,\cdots,r). \tag{2.9}$$

(iii) If $G = \Psi_\alpha$, let ν be a measure on $\mathcal{B}(-\infty, 0])$ determined by $\nu(x, \infty] = (-x)^\alpha$, $x < 0$, and T be a PRM(ν) on $(0, \infty]$, and let $H(x) = \exp\{-T((x, 0])\}$, then for any $x_i \le 0$, $i = 1, \cdots, r$,

$$(H_{n,n}(x_i, \cdot), i = 1, 2, \cdots, r) \xrightarrow{d} (H(x_i, \cdot), i = 1, 2, \cdots, r). \tag{2.10}$$

Proof. For (i), we can write

$$T_{n,n} = \sum_{k=1}^{n} \epsilon_{a_n^{-1}(X_k - b_n)}$$

where ϵ_a is the delta measure at a. Then, by Corollary 4.19 of Resnick (1987), T_n converges weakly to a $PRM(\mu)$. Therefore the continuous mapping theorem gives the result. Proofs for (ii) and (iii) are similar.□

Remark 2.1 Modifying the Proof of Theorem 2.4, it can be shown that if $m \sim d \cdot n$ for some $0 < d \le 1$, then $H_{n,m}(x, \cdot) \xrightarrow{d} \exp(-d \cdot V_d(x, \cdot))$ where $V_d(x, \omega)$ is a Poisson random variable with the mean $d^{-1}c(x)$. Therefore the bootstrap is inconsistent if m has the same order of n.

2.3 Bootstrapping Extremes when m=o(n)

In this section it will be shown that bootstrap distribution can be made consistent by choosing the resample size suitably. Swanepoel (1986) proved that the bootstrap is strongly consistent for the maximum of uniform random variables, if $m_n = o(n^{(1+\epsilon)/2}(\log n)^{-1/2})$. Deheuvels, Mason and Schorack (1993) proved the weak and strong consistency of the bootstrap distribution of the maximum with suitable choices of resample sizes. They utilized a unified form of extreme value distribution (von Mises parameterization), and thus the bootstrap distribution they define does not depend on the domain of attraction to which F belongs. In contrast, we define different bootstrap distributions for each of three types of domains of attraction and prove their consistency for appropriate choices of resample sizes. An advantage of our approach will become clear when results in this chapter are applied to the inference for population parameters in Chapter 4.

2.3.1 Known normalizing constants

The next lemma is important in its own right.

Lemma 2.1 *Suppose that $\{H_n(\cdot,\omega), n = 1, 2, \cdots, \infty\}$ is a sequence of random cdfs on a probability space $(\Omega, \mathcal{F}, P)$ such that there exists a countable dense set $D \subset \mathbf{R}$ and for each $x \in D$,*

$$H_n(x,\cdot) \to H_\infty(x,\cdot), \tag{2.11}$$

w.p.1 . Then

$$\rho(H_n, H_\infty) \to 0, \tag{2.12}$$

w.p.1, where ρ is a metric on the space of cdf's which is equivalent to weak convergence. In particular, if H_∞ is continuous w.p.1,

$$\sup_{x\in\mathbf{R}} \mid H_n(x,\cdot) - H_\infty(x,\cdot) \mid \to 0, \tag{2.13}$$

w.p.1 . If (2.11) holds in probability, then so do (2.12) and (2.13).

Proof. Let (2.11) hold w.p.1 . Note that the exceptional set may depend on $x \in D$. Let $\{r_j\}$ be an enumeration of D. For each $r_j \in D$, there exists $A_j \in \mathcal{F}$ such that $P(A_j) = 1$ and for each $\omega \in A_j$, $H_n(r_j,\omega) \to H_\infty(r_j,\omega)$. Let $A = \cap_{j=1}^{\infty} A_j$. Then $P(A) = 1$ and for each $\omega \in A$ and $r_j \in D$,

$$H_n(r_j,\omega) \to H_\infty(r_j,\omega).$$

Thus for fixed $\omega \in A$, $H_n(\cdot,\omega)$ converges weakly to $H_\infty(\cdot,\omega)$. Therefore for each $\omega \in A$,

$$\rho(H_n(\cdot,\omega), H_\infty(\cdot,\omega)) \to 0,$$

which proves (2.12).

If $H_\infty(\cdot,\omega)$ is continuous for all $\omega \in B$ such that $P(B) = 1$, then by Polya's theorem,

$$\sup_{x\in\mathbf{R}} \mid H_n(x,\omega) - H_\infty(x,\omega) \mid \to 0,$$

for each $\omega \in A \cap B$. This proves (2.13).

Now let (2.11) hold in probability. Let $\{r_j\}$ be an enumeration of D. For any subsequence $\{n_i^{(0)}\}_{i=1}^\infty$ of $\{n\}_{n=1}^\infty$, there exists a further subsequence $\{n_i^{(1)}\}_{i=1}^\infty$ of $\{n_i^{(0)}\}_{i=1}^\infty$ such that

$$H_{n_i^{(1)}}(r_1,\cdot) \to H_\infty(r_1,\cdot), \quad w.p.1,$$

as $i \to \infty$. By repeating this procedure, we have a sequence $\{C_j\}$ of elements of $\mathcal{F}$ and an array $\{n_i^{(j)}; i = 1,2,\cdots\}_{j=1}^\infty$ of subsequences of $\{n\}_{n=1}^\infty$ such that for every $j < k$, $\{n_i^{(k)}\}_{i=1}^\infty$ is a subsequence of $\{n_i^{(j)}\}_{i=1}^\infty$ and for each j, $P(C_j) = 1$ and for each $j \in \mathbf{N}$ and each $\omega \in C_j$,

$$H_{n_i^{(j)}}(r_j,\omega) \to H_\infty(r_j,\omega),$$

as $i \to \infty$. Now define a new sequence $\{n_i^*\}_{i=1}^\infty$ by $n_i^* = n_i^{(i)}$ for each i. Let $C = \cap_{j=1}^\infty C_j$. Then $P(C) = 1$ and for each $\omega \in C$ and $r_j \in D$,

$$H_{n_i^*}(r_j,\omega) \to H_\infty(r_j,\omega),$$

as $i \to \infty$. Therefore

$$\rho(H_{n_i^*}, H_\infty) \to 0,$$

w.p.1 . This is equivalent to $\rho(H_n, H_\infty) \xrightarrow{p} 0$ (e.g., Chow and Teicher, 1985, pp. 67), where $\xrightarrow{p}$ denotes the convergence in probability. If $H_\infty(\cdot,\omega)$ is continuous for all $\omega \in E$ such that $P(E) = 1$, then by Polya's theorem (e.g., Chow and Teicher, 1985, pp. 265),

$$\sup_{x\in\mathbf{R}} \mid H_{n_i^*}(x,\omega) - H_\infty(x,\omega) \mid \rightarrow 0,$$

for each $\omega \in C \cap E$, which yields (2.13) in probability.

The following is an alternative Proof of the convergence in probability. Let

$$W_n(\omega) = \sum_{j=1}^{\infty} \mid H_n(r_j,\omega) - H_\infty(r_j,\omega) \mid 2^{-j}$$

where $\{r_j\}$ is an enumeration of D. For each $\epsilon > 0$, there exists N_ϵ such that $\sum_{j=N_\epsilon+1}^{\infty} 2^{-j+1} < \frac{\epsilon}{2}$. Therefore

$$\begin{aligned} P\{W_n(\omega) > \epsilon\} &\leq P\{\sum_{j=1}^{N_\epsilon} \mid H_n(r_j,\omega) - H_\infty(r_j,\omega) \mid 2^{-j} > \frac{\epsilon}{2}\} \\ &\rightarrow 0, \end{aligned}$$

by hypotheses. Therefore $W_n \xrightarrow{p} 0$. Thus, for any subsequence $\{n'\}$ of $\{n\}$, there exists a further subsequence $\{n''\}$ of $\{n'\}$ and $A \in \mathcal{F}$ such that $P(A) = 1$ and for each $\omega \in A$, $W_{n''}(\omega) \rightarrow 0$. So for each $\omega \in A$ and each $r_j \in$D, $H_{n''}(r_j,\omega) \rightarrow H_\infty(r_j,\omega)$. Therefore for each $\omega \in A$,

$$\rho(H_{n''}(\cdot,\omega), H_\infty(\cdot,\omega)) \rightarrow 0,$$

(e.g., Chow and Teicher, 1985, pp. 257). Hence by the choice of n'', $\rho(H_n(\cdot,\omega), H_\infty(\cdot,\omega)) \xrightarrow{p} 0$. If $H_\infty(\cdot,\omega)$ is continuous for all $\omega \in B$ such that $P(B) = 1$, then by Polya's theorem,

$$\sup_{x\in\mathbf{R}} \mid H_{n''}(x,\omega) - H_\infty(x,\omega) \mid \xrightarrow{p} 0,$$

for each $\omega \in A \cap B$, which yields convergence in probability result.□

The next theorem shows that if m increases to infinity but slower than n, the bootstrap distribution is consistent.

Theorem 2.5 *Suppose that (2.1) holds. If m=o(n),*

$$\sup_{x \in \mathbf{R}} \mid H_{n,m}(x,\cdot) - G(x) \mid \to 0 \tag{2.14}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for each* $\lambda \in (0,1)$ *then (2.14) is true w.p.1.*

Proof. First we will prove the convergence in probability result. Let $p_m = 1 - F(a_m x + b_m)$ and $c(\cdot)$ as in (2.5). Since

$$\begin{aligned} m(1 - F_n(a_m x + b_m)) &= \frac{m}{n} n(1 - F_n(a_m x + b_m)) \\ &= \frac{m}{n} T_{n,m}, \end{aligned}$$

$E(\frac{m}{n} T_{n,m}) = m p_m \to c(x)$, and $Var(\frac{m}{n} T_{n,m}) = \frac{m}{n} m p_m (1 - p_m) \to 0$, it follows that $m(1 - F_n(a_m x + b_m)) \xrightarrow{P} c(x)$ for each $x \in \mathbf{R}$. Therefore, from (2.6),

$$H_{n,m}(x,\cdot) \xrightarrow{P} e^{-c(x)} = G(x)$$

for each $x \in \mathbf{R}$. Now Lemma 2.1 yields (2.14) in probability.

To prove the convergence w.p.1 result, since $\frac{m}{n} T_{n,m} = \frac{m}{n}(T_{n,m} - n p_m) + m p_m$ and $m p_m \to c(x)$, we need to show that only $\frac{m}{n}(T_{n,m} - n p_m) \to 0$ w.p.1 to prove that $\frac{m}{n} T_{n,m} \to c(x)$ w.p.1. By the Borel-Cantelli lemma, it is enough to show that for each $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P\{\mid \frac{m}{n}(T_{n,m} - n p_m) \mid > \varepsilon\} < \infty.$$

Let $\varphi_m(\theta) = p_m e^\theta + (1 - p_m)$ be the moment generating function of Bernoulli distribution with the parameter p_m. Then for each $\theta > 0$,

$$\begin{aligned}
\frac{m}{n}\log P\{\frac{m}{n}(T_{n,m} - np_m) > \varepsilon\} &= \frac{m}{n}\log P(e^{\theta(T_{n,m}-np_m)} > e^{\theta\frac{n}{m}\varepsilon}) \qquad (2.15)\\
&\leq \frac{m}{n}\log(\{\exp(-\theta\frac{n}{m}\varepsilon)\}E\{\exp\theta(T_{n,m} - np_m)\})\\
&= -\theta\varepsilon - \theta m p_m + \log\varphi_m(\theta)^m\\
&\rightarrow -\theta\varepsilon - \theta c(x) + \log\{\exp(c(x)(e^\theta - 1))\}\\
&= -\theta\varepsilon + c(x)(e^\theta - 1 - \theta)\\
&= f(\theta, \varepsilon) \quad (say)
\end{aligned}$$

By taking the derivative of $f(\theta,\varepsilon)$, we can show that $\theta_0(n) := \log(\frac{c+\varepsilon}{c})$ minimizes $f(\theta,\varepsilon)$, where $c = c(x)$. Now let

$$g(\varepsilon) := f(\theta_0, \varepsilon) = -(\varepsilon + c)\log(\frac{c+\varepsilon}{c}) + \varepsilon,$$

then

$$g(0) = 0$$

and

$$\begin{aligned}
g'(\varepsilon) &= 1 - \log(\frac{c+\varepsilon}{c}) - (\varepsilon + c)\frac{1}{\frac{c+\varepsilon}{c}}\frac{1}{c}\\
&= -\log(\frac{c+\varepsilon}{c}) < 0,
\end{aligned}$$

for each $\varepsilon > 0$. Thus $g(\varepsilon) < 0$ for each $\varepsilon > 0$. Define

$$g_n(\varepsilon) := \frac{m}{n}\log\{\exp(-\theta_0(\varepsilon)\frac{n}{m}\varepsilon)\}E\{\exp(\theta_0(\varepsilon)(T_{n,m} - np_m))\}$$

then $g_n(\varepsilon) \rightarrow g(\varepsilon)$. Let $\theta = \theta_0(\varepsilon)$ in (2.15), then

$$\sum_{n=1}^{\infty} P\{\frac{m}{n}(T_{n,m} - np_m) > \varepsilon\} = \sum_{n=1}^{\infty}\exp\{\log P(\frac{m}{n}(T_{n,m} - np_m) > \varepsilon)\}$$

$$\leq \sum_{n=1}^{\infty} \exp\{\frac{n}{m} g_n(\varepsilon)\}.$$

Now, given $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that $g(\varepsilon) + \delta_\varepsilon < 0$ and there exists $N_\varepsilon \in \mathbf{N}$ such that $g_n(\varepsilon) < g(\varepsilon) + \delta_\varepsilon$ for every $n \geq N_\varepsilon$.

Therefore

$$\begin{aligned}
\sum_{n=1}^{\infty} \exp\{\frac{n}{m} g_n(\varepsilon)\} &= \sum_{n=1}^{N_\varepsilon - 1} \exp\{\frac{n}{m} g_n(\varepsilon)\} + \sum_{n=N_\varepsilon}^{\infty} \exp\{\frac{n}{m} g_n(\varepsilon)\} \\
&\leq \sum_{n=1}^{N_\varepsilon - 1} \exp\{\frac{n}{m} g_n(\varepsilon)\} + \sum_{n=N_\varepsilon}^{\infty} \exp\{\frac{n}{m} (g(\varepsilon)\} + \delta_\varepsilon) \\
&< \infty \quad (\textit{by the assumption}).
\end{aligned}$$

Hence $\sum_{n=1}^{\infty} P\{\frac{m}{n}(T_{n,m} - np_m) > \varepsilon\} < \infty$ for each $\varepsilon > 0$. By similar reasoning we can show that $\sum_{n=1}^{\infty} P\{\frac{m}{n}(T_{n,m} - np_m) < -\varepsilon\} < \infty$ for each $\varepsilon > 0$. □

2.3.2 Unknown normalizing constants

If F is unknown, a_m and b_m need to be estimated from the data for $H_{n,m}(\cdot)$ to be of use. Let $\hat{a}_m$ and $\hat{b}_m$ be some estimators of a_m and b_m based on $X_1, X_2, \cdots, X_n$. Now, define the bootstrap distribution of $\dot{a}_n^{-1}(X_{n:n} - b_n)$ with estimated normalizing constants by

$$\hat{H}_{n,m}(x, \omega) = P\{\hat{a}_m^{-1}(Y_{m:m} - \hat{b}_m) \leq x | \mathbf{X}_n\}.$$

The next theorem gives a sufficient condition for $\hat{H}_{n,m}(x, \omega)$ to be consistent.

Theorem 2.6 *Assume that $\hat{a}_m$, $\hat{b}_m$, and m_n satisfy the following:*

(*i*) *There exists $D \subset \mathbf{R}$ such that for each $x \in D$,*

$H_{n,m}(x, \cdot) \to G(x)$, *w.p.1*,

$$(ii) \quad \frac{\hat{a}_m}{a_m} \to 1, \ w.p.1,$$

$$(iii) \quad a_m^{-1}(\hat{b}_m - b_m) \to 0, \ w.p.1.$$

Then

$$\sup_{x \in \mathbf{R}} \mid \hat{H}_{n,m}(x, \cdot) - G(x) \mid \to 0, \ w.p.1. \tag{2.16}$$

Theorem also holds if "w.p.1" is replaced by "in probability".

Proof. (Step 1.) Suppose that (i), (ii) and (iii) hold. Fix $x > 0$. Then (i) is equivalent to the existence of $\Omega_1 \in \mathcal{F}$ such that $P(\Omega_1) = 1$ and for all $\omega \in \Omega_1$, $m\{1 - F_n(a_m x + b_m)\} \to c(x)$. From (ii) and (iii), there exist $\Omega_2 \in \mathcal{F}$ such that $P(\Omega_2) = 1$ and $\hat{a}_m(\omega)/a_m \to 1$ and $a_m^{-1}(\hat{b}_m(\omega) - b_m) \to 0$ for every $\omega \in \Omega_2$. So, for every $\omega \in \Omega_1 \cap \Omega_2$ and $\varepsilon > 0$, there exists $N(\varepsilon, \omega) \in \mathbf{N}$ such that for every $n \geq N(\varepsilon, \omega)$, $(1 - \varepsilon)a_m < \hat{a}_m(\omega) < (1 + \varepsilon)a_m$ and $-\varepsilon a_m < \hat{b}_m(\omega) - b_m < \varepsilon a_m$. Thus for $x > 0$ and $\omega \in \Omega_1 \cap \Omega_2$,

$$m\{1 - F_n(\hat{a}_m(\omega)x + \hat{b}_m(\omega))\} \leq m\{1 - F_n(((1 - \varepsilon)x - \varepsilon)a_m + b_m),$$

and

$$\overline{\lim}_{n \to \infty} m\{1 - F_n(\hat{a}_m(\omega)x + \hat{b}_m(\omega))\} \leq c((1 - \varepsilon)x - \varepsilon).$$

A similar inequality holds for lim inf. Since $c(x)$ is continuous, for $\omega \in \Omega_1 \cap \Omega_2$,

$$\lim_{n \to \infty} m\{1 - F_n(\hat{a}_m(\omega)x + \hat{b}_m(\omega))\} = c(x).$$

A similar argument holds for $x < 0$. Therefore we have shown that for every $x \in D$, $\hat{H}_{n,m}(x, \omega) \to G(x)$ w.p.1 . Now Lemma 2.1 applies, to obtain (2.16).

(Step 2.) Now suppose that (i), (ii) and (iii) hold in probability. Then arguing as in the Proof of Lemma 1, for any subsequence $\{n_i\}_{i=1}^{\infty}$ of $\{n\}_{n=1}^{\infty}$ there exists a further subsequence $\{n_i^*\}_{i=1}^{\infty}$ of $\{n_i\}_{i=1}^{\infty}$ such that (i), (ii) and (iii) hold w.p.1 for each $x \in D$. Now by Step 1,

$$\sup_{x\in\mathbf{R}} \mid \hat{H}_{n_i^*,m(n_i^*)}(x,\cdot) - G(x) \mid \rightarrow 0,$$

w.p.1 . Thus (2.16) holds in probability. □

For the bootstrap distribution to be consistent, as is seen from Theorem 2.5 and 2.6, we need to choose $\hat{a}_m$ and $\hat{b}_m$ satisfying (ii) and (iii) in Theorem 2.6. Since a_m and b_m are functionals of F, natural choices of $\hat{a}_m$ and $\hat{b}_m$ are the empirical counter parts of a_m and b_m. We define $\hat{a}_m$ and $\hat{b}_m$ for each domain of attraction. Let $l_n = [\frac{n}{m}]$ and $l_n' = [\frac{n}{em}]$, where $[\cdot]$ is the integer part.

Theorem 2.7 *Define*

$$\begin{aligned}
&\textit{(i) if } F \in D(\Lambda), & \hat{a}_m &= F_n^{-1}(1-\frac{1}{em}) - F_n^{-1}(1-\frac{1}{m}) = X_{n-l_n':n} - X_{n-l_n:n},\\
& & \hat{b}_m &= F_n^{-1}(1-\frac{1}{m}) = X_{n-l_n:n},\\
&\textit{(ii) if } F \in D(\Phi_\alpha), & \hat{a}_m &= F_n^{-1}(1-\frac{1}{m}) = X_{n-l_n:n},\\
& & \hat{b}_m &= 0,\\
&\textit{(iii) if } F \in D(\Psi_\alpha), & \hat{a}_m &= \theta_{F_n} - F_n^{-1}(1-\frac{1}{m}) = X_{n:n} - X_{n-l_n:n},\\
& & \hat{b}_m &= \theta_{F_n} = X_{n:n}.
\end{aligned}$$

If m=o(n), then

$$\sup_{x\in\mathbf{R}} \mid \hat{H}_{n,m}(x,\cdot) - G(x) \mid \rightarrow 0 \tag{2.17}$$

in probability where $G = \Lambda, \Phi_\alpha$ or Ψ_α according to the domain of attraction of F. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $\lambda \in (0,1)$ then (2.17) is true w.p.1.

Remark 2.2 Combining (2.1) and (2.17) we see that

$$\sup_{x\in\mathbf{R}} \mid \hat{H}_{n,m}(x,\cdot) - G_n(x) \mid \to 0, \tag{2.18}$$

in probability if m=o(n), and w.p.1 if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $0 < \lambda < 1$, where G_n is as in (2.3). Therefore $\hat{H}_{n,m}(x,\omega)$ approximates $G_n(x)$ uniformly in $\mathbf{R}$ when $n \to \infty$. Note also that $m = o(n/\log n)$ is sufficient for $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for every $\lambda \in (0,1)$. Deheuvels, Mason and Schrack (1993) proved the strong consistency under the weaker condition $m = o(n(\log\log n)^{-1})$.

Proof of Theorem 2.7. We provide the Proof only for $F \in D(\Psi_\alpha)$. From Theorem 2.6, it suffices to show that

$$\frac{\hat{a}_m}{a_m} = \frac{X_{n:n} - X_{n-l_n:n}}{\theta_F - \gamma_m} \to 1$$

and

$$a_m^{-1}(\hat{b}_m - b_m) = \frac{X_{n:n} - \theta_F}{\theta_F - \gamma_m} \to 0,$$

both in probability or w.p.1 . Since

$$\frac{X_{n:n} - X_{n-l_n:n}}{\theta_F - \gamma_m} = \frac{X_{n:n} - \theta_F}{\theta_F - \gamma_n} \cdot \frac{\theta_F - \gamma_n}{\theta_F - \gamma_m} - \frac{X_{n-l_n:n} - \theta_F}{\theta_F - \gamma_m},$$

it is sufficient to show that

$$\frac{\theta_F - \gamma_n}{\theta_F - \gamma_m} \to 0 \tag{2.19}$$

and

$$\frac{X_{n-l_n:n} - \theta_F}{\theta_F - \gamma_m} \to -1, \tag{2.20}$$

in probability or w.p.1.

We prove (2.19) first. Since $F \in D(\Psi_\alpha)$, (2.2) holds. (2.2) is equivalent to $1 - F(\theta_F - t) = t^\alpha L(t)$, where L is a slowly varying function at 0. Under this it is known that $1 - n^{-1} \sim F(\gamma_n)$ (cf. Leadbetter *et. al*, 1983, pp. 18). Therefore, as $n \to \infty$,

$$\frac{m}{n} \sim \frac{1 - F(\gamma_n)}{1 - F(\gamma_m)} = (\frac{\theta_F - \gamma_n}{\theta_F - \gamma_m})^\alpha \frac{L(\theta_F - \gamma_n)}{L(\theta_F - \gamma_m).} \tag{2.21}$$

Now suppose $(\theta_F - \gamma_n)/(\theta_F - \gamma_m)$ does not converge to 0. Then there exist a subsequence $\{n_i\}_{i=1}^\infty$ of $\{n\}_{n=1}^\infty$ and a constant c, $0 < c \leq 1$ such that $(\theta_F - \gamma_{n_i})/(\theta_F - \gamma_{m(n_i)}) \to c$, as $i \to \infty$. Since $L(tx)/L(t) \to 1$, as $t \to 0$ locally uniformly in $x > 0$ (cf. Seneta, 1976, Theorem 1.1),

$$\frac{L(\theta_F - \gamma_{n_i})}{L(\theta_F - \gamma_{m(n_i)})} = \frac{L\{(\theta_F - \gamma_{n_i})(\theta_F - \gamma_{m(n_i)})^{-1}(\theta_F - \gamma_{m(n_i)})\}}{L(\theta_F - \gamma_{m(n_i)})} \to 1,$$

as $i \to \infty$, and thus $\dfrac{1 - F(\gamma_{n_i})}{1 - F(\gamma_{m(n_i)})} \to c^\alpha > 0$ which contradicts (2.21), since $m/n \to 0$. Hence (2.19) is proved.

To prove (2.20), for arbitrary $0 < \varepsilon < 1$, let $u_m(\varepsilon) = \theta_F - (1 - \varepsilon)(\theta_F - \gamma_m)$ and $S_{n,m}(x) := \#\{i : 1 \leq i \leq n, X_i > x\}$. Then

$$\begin{aligned} P(\frac{X_{n-l_n:n} - \theta_F}{\theta_F - \gamma_m} - (-1) > \varepsilon) &= P(X_{n-l_n:n} > u_m(\varepsilon)) \\ &= P(S_{n,m}(u_m(\varepsilon)) > l_n) \\ &\sim P(\frac{m}{n}\sum_{i=1}^n I(X_i > u_m) > 1) \\ &= P\{m(1 - F(u_m(\varepsilon))) > 1\} \qquad (2.22) \\ &\to 0, \end{aligned}$$

since $m\{1 - F_n(u_m(\varepsilon))\} \xrightarrow{P} c(\varepsilon - 1) = (1 - \varepsilon)^\alpha < 1$ as was proved in Theorem 2.5. Next, let $u'_m(\varepsilon) = \theta_F - (1 + \varepsilon)(\theta_F - \gamma_m)$. Similarly,

$$P(\frac{X_{n-l_n:n} - \theta_F}{\theta_F - \gamma_m} - (-1) < -\varepsilon) = P(S_{n,m}(u'_m(\varepsilon) \leq l_n)$$

$$\begin{aligned} &\sim\quad P(\frac{m}{n}\sum_{i=1}^{n} I(X_i > u'_m(\varepsilon) \le 1) \\ &=\quad P\{m(1-F(u'_m(\varepsilon))) \le 1\} \\ &\to\quad 0, \end{aligned}$$

since $m\{1 - F_n(u_m(\varepsilon))\} \xrightarrow{P} c(\varepsilon + 1) = (1+\varepsilon)^\alpha > 1$. Hence $\frac{X_{n-l_n:n}-\theta_F}{\theta_F-\gamma_m} \xrightarrow{P} -1$.

Now suppose that $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $0 < \lambda < 1$. Then from (2.22),

$$\begin{aligned} \sum_n^{\infty} P(\frac{X_{n-l_n:n}-\theta_F}{\theta_F-\gamma_m} - (-1) > \varepsilon) &\sim \sum_n^{\infty} P\{m(1-F(u_m(\varepsilon))) > 1\} \\ &< \infty, \end{aligned}$$

since $m\{1 - F_n(u_m(\varepsilon))\} \to c(\varepsilon - 1) = (1-\varepsilon)^\alpha$ w.p.1. Hence, by Borel-Cantelli's lemma, $\frac{X_{n-l_n:n}-\theta_F}{\theta_F-\gamma_m} \to -1$ w.p.1. □

Remark 2.3 Note that (2.19) does not involve any random variables. Note also that, from the proof above, $\frac{X_{n-l_n:n}-\theta_F}{\theta_F-\gamma_m} \to -1$ in probability (w.p.1) whenever $m\{1 - F_n(a_m x + b_m)\} \to c(x)$ in probability (w.p.1) for each $x \in \mathbf{R}$. Thus the method of the above proof can be employed in the Proof of the corresponding theorem for stationary processes.

The following theorem shows that the joint distribution of $a_n^{-1}(X_{n:n} - b_n), a_n^{-1}(X_{n-1:n} - b_n), \cdots, a_n^{-1}(X_{n-r+1:n} - b_n)$ can be bootstrapped consistently. Let

$$F_r(x_1, \cdots, x_r) \equiv \lim_{n\to\infty} P\{a_n^{-1}(X_{n:n} - b_n) \le x_1, \cdots, a_n^{-1}(X_{n-r+1:n} - b_n) \le x_r\}.$$

Theorem 2.8 *Assume the hypotheses on F and choose $\hat{a}_m$ and $\hat{b}_m$ as in Theorem 2.7 according to the domain of attraction of F. If $m=o(n)$, then*

$$\sup_{x_1>\cdots>x_r} \mid P\{\hat{a}_m^{-1}(Y_{m:m}-\hat{b}_m)\leq x_1,\cdots,\hat{a}_m^{-1}(Y_{m-r+1:m}-\hat{b}_m)\leq x_r|\mathbf{X}_n\}$$

$$-F_r(x_1,\cdots,x_r)\mid\to 0, \tag{2.23}$$

in probability. Moreover, if $\sum_{n=1}^{\infty}\lambda^{\frac{n}{m}}<\infty$ for each $\lambda\in(0,1)$ then (2.23) is true w.p.1.

Proof. For $x_1<x_2<\cdots<x_r$, define $u_n^{(i)}=a_nx_i+b_n$ and $S_n^{(i)}(x)=\#\{j:1\leq j\leq n, X_j>u_n^{(i)}\}$. Define also

$$K\equiv\{(k_1,\cdots,k_{r-1}):k_i\geq 0, i=1,\cdots,r-1, k_1+k_2+\cdots+k_j\leq j, j=1,2,\cdots,r-1\}.$$

Then

$$\begin{aligned}
& P\{a_n^{-1}(X_{n:n}-b_n)\leq x_1, a_n^{-1}(X_{n-1:n}-b_n)\leq x_2,\cdots,a_n^{-1}(X_{n-r+1:n}-b_n)\leq x_r\}\\
= \; & P\{X_{n:n}\leq u_n^{(1)}, X_{n-1:n}\leq u_n^{(2)}\cdots, X_{n-r+1:n}\leq u_n^{(r)}\}\\
= \; & P\{S_n^{(1)}=0, S_n^{(2)}\leq 1,\cdots,S_n^{(r)}\leq r-1\}\\
= \; & \sum_{(k_1,\cdots,k_{r-1})\in K} P\{S_n^{(1)}=0, S_n^{(2)}-S_n^{(1)}=k_1,\cdots,S_n^{(r)}-S_n^{(r-1)}=k_{r-1}\}\\
= \; & \sum_{(k_1,\cdots,k_{r-1})\in K}\binom{n}{k_1}(F(u_n^{(1)})-F(u_n^{(2)}))^{k_1}\cdots\\
& \binom{n-k_1-k_2\cdots-k_{r-2}}{k_{r-1}}(F(u_n^{(r-1)})-F(u_n^{(r)}))^{k_{r-1}}(F(u_n^{(r)}))^{n-\sum_1^{r-1}k_i}\\
\to \; & \sum_{(k_1,\cdots,k_{r-1})\in K}\frac{\{\log(G(x_2))-\log(G(x_1))\}^{k_1}}{k_1!}\cdots\\
& \frac{\{\log(G(x_r))-\log(G(x_{r-1}))\}^{k_{r-1}}}{k_{r-1}!}(G(x_r)).
\end{aligned} \tag{2.24}$$

Let $\hat{u}_m^{(i)} = \hat{a}_n x_i + \hat{b}_n$, then

$$
\begin{aligned}
& P\{\hat{a}_m^{-1}(Y_{m:m} - \hat{b}_m) \le x_1, Y_{m-1:m} - \hat{b}_m) \le x_2 \cdots, \hat{a}_m^{-1}(Y_{m-r+1:m} - \hat{b}_m) \le x_r | \mathbf{X}_n\} \\
= & \sum_{(k_1,\cdots,k_{r-1})\in K} \binom{n}{k_1} (F_n(u_m^{(1)}) - F_n(u_m^{(2)}))^{k_1} \cdots \\
& \binom{n - k_1 - k_2 \cdots - k_{r-2}}{k_{r-1}} (F_n(u_m^{(r-1)}) - F_n(u_m^{(r)}))^{k_{r-1}} (F_n(u_m^{(r)}))^{n - \sum_1^{r-1} k_i}
\end{aligned}
$$

which converges to (2.24) in probability if $m/n \to 0$ and w.p.1 if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $\lambda \in (0,1)$.□

Theorem 2.8 and the continuous mapping theorem give the following.

Theorem 2.9 *Assume the hypothesis on F and choose $\hat{a}_m$ and $\hat{b}_m$ as in Theorem 2.7 according to the domain of attraction of F. Let $f : \mathbf{R}^r \to \mathbf{R}^l$ be continuous a.e. with respect to $F_r(\cdot, \cdots, \cdot)$. If $m=o(n)$, then*

$$
\begin{aligned}
\sup_{\mathbf{y}\in\mathbf{R}^l} & \mid P\{f(\hat{a}_m^{-1}(Y_{m:m} - \hat{b}_m), \hat{a}_m^{-1}(Y_{m-1:m} - \hat{b}_m), \cdots, \hat{a}_m^{-1}(Y_{m-r+1:m} - \hat{b}_m)) \le \mathbf{y} \mid \mathbf{X}_n\} \\
& - P\{f(a_n^{-1}(X_{n:n} - b_n), a_n^{-1}(X_{n-1:n} - b_n) \cdots, a_n^{-1}(X_{n-r+1:n} - b_n)) \le \mathbf{y}\} \mid \to 0 \quad (2.25)
\end{aligned}
$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $\lambda \in (0,1)$ then (2.25) is true w.p.1.

2.4 Resampling from the Smoothed Empirical Distribution

Instead of resampling from the empirical distribution F_n, one can think of resampling from a smoothed empirical distribution. Especially, the bootstrap using kernel estimators of F has been investigated by Efron (1982), Silverman and Young (1987), Hall, DiCiccio and Romano (1989) and Falk and Reiss (1989). The smoothed bootstrap described in this section will be used in Chapter 4 where it is necessary to use a continuous cdf in the resampling scheme. The kernel estimator $\tilde{F}_n$ of F is defined as

$$\tilde{F}_n(x) = n^{-1}\sum_{i=1}^{n} K((x - X_i)/h_n), \tag{2.26}$$

where $h_n = h_n(X_1, \cdots, X_n) > 0$ and $h_n : \mathbf{R}^n \to \mathbf{R}^+$ is Borel measurable and $K : \mathbf{R} \to [0, 1]$ is a continuous cdf . Given $\mathbf{X}_n$, let $\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_m$ be conditionally i.i.d. random variables having the cdf $\tilde{F}_n$. Define the smoothed bootstrap distribution of $a_n^{-1}(X_{n:n} - b_n)$ by

$$\tilde{H}_{n,m}(x, \omega) = P(a_m^{-1}(\tilde{Y}_{m:m} - b_m) \leq x|\mathbf{X}_n).$$

Define for $z \in \mathbf{R}$,

$$\tilde{D}_n(z, \omega) = m\{1 - F_n(a_m(x - a_m^{-1}h_n z) + b_m)\}.$$

Before proving the consistency result we need the following.

Lemma 2.2 *Suppose that (2.1) holds and assume that*

$$(i) \quad m(1 - F_n(a_m x + b_m)) \to c(x), \ w.p.1,$$

$$(ii) \quad a_m^{-1}h_n \to 0, \ w.p.1.$$

Then for each $z \in \mathbf{R}$,

$$\tilde{D}_n(z, \cdot) \to c(x), \ w.p.1. \tag{2.27}$$

If (i) and (ii) hold in probability, so does (2.27).

Proof. (Step 1) Suppose that (i) and (ii) hold. Since $m(1 - F_n(a_m x + b_m))$ is nonincreasing in x and $c(x)$ is continuous, the convergence in (i) is locally uniform, i.e., for each $M > 0$ there exists $A \in \mathcal{F}$ such that $P(A) = 1$ and for each $\omega \in A$,

$$\sup_{y:|y-x|\leq M} | m(1 - F_n(a_m y + b_m)) - c(y) | \to 0.$$

From (ii), there exists $B \in \mathcal{F}$ such that $P(B) = 1$ and for each $\omega \in B$, $a_m^{-1} h_n \to 0$. Therefore for each $\omega \in A \cap B$ and large enough n,

$$\begin{aligned}
& | \tilde{D}_n(z,\omega) - c(x) | \\
\leq\ & | m\{1 - F_n(a_m(x - a_m^{-1}h_n z) + b_m)\} - c(x - a_m^{-1}h_n z) | \\
& + | c(x - a_m^{-1}h_n z) - c(x) | \\
\leq\ & \sup_{y:|y-x|\leq M} | m(1 - F_n(a_m y + b_m)) - c(y) | + | c(x - a_m^{-1}h_n z) - c(x) | \\
\to\ & 0.
\end{aligned}$$

This proves (2.27).

(Step 2) Now suppose (i) and (ii) hold in probability. For any subsequence $\{n_i^{(0)}\}_{i=1}^{\infty}$ of $\{n\}_{n=1}^{\infty}$, there exists a further subsequence $\{n_i^{(1)}\}_{i=1}^{\infty}$ of $\{n_i^{(0)}\}_{i=1}^{\infty}$ such that along $\{n_i^{(1)}\}_{i=1}^{\infty}$ (i) and (ii) hold w.p.1. By the Step 1, along $\{n_i^{(1)}\}_{i=1}^{\infty}$ (2.27) holds w.p.1. for each $z \in \mathbf{R}$. This implies (2.27) in probability. □

We have consistency results of $\tilde{H}_{n,m}$ under some conditions on K and h_n.

Theorem 2.10 *Suppose that (2.1) holds and the support of K is bounded above. Assume:*

(*i*) *There exists a countable dense set $D \subset \mathbf{R}$ such that for each $x \in D$,*

$$m(1 - F_n(a_m x + b_m)) \to c(x), \quad w.p.1,$$

(*ii*) $a_m^{-1} h_n \to 0, \quad w.p.1.$

Then

$$\sup_{x \in \mathbf{R}} \mid \tilde{H}_{n,m}(x, \cdot) - G(x) \mid \to 0 \tag{2.28}$$

w.p.1. If (i) and (ii) hold in probability then so does (2.28).

Remark 2.4 From Theorem 2.5, the condition $m = o(n)$ and the condition $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for every $\lambda \in (0, 1)$ are sufficient for (i) to hold in probability and w.p.1. respectively.

Remark 2.5 It can be easily seen that Theorem 2.6 holds when $Y_{m:m}$ is replaced by $\tilde{Y}_{m:m}$ and therefore Theorem 2.10 still holds when a_m and b_m are replaced by $\hat{a}_m$ and $\hat{b}_m$ for which $\hat{a}_m/a_m \to 1$ and $a_m^{-1}(\hat{b}_m - b_m) \to 0$. The smoothed bootstrap version of Theorem 2.9 also holds.

Remark 2.6 Any upper endpoint U of the support of K yields the same asymptotic result in Theorem 2.10. But for $F \in D(\Psi_\alpha)$ a reasonable choice is $U = 0$ because if $U > 0$ the bootstrap distribution $P\{(\tilde{Y}_{m:m} - X_{n:n})/(X_{n:n} - X_{n-l_n:n}) \leq x \mid \mathbf{X}_n\}$ will have a positive mass on the positive real line while $P\{(X_{n:n} - \theta_F)/(\theta_F - \gamma_n) \leq x\}$ does not.

Proof of Theorem 2.10. Since $\tilde{H}_{n,m}(x, \omega) = \tilde{F}_n^m(a_m x + b_m)$, it is enough to show that $D_n \equiv mP\{\tilde{Y}_1 > (a_m x + b_m) \mid \mathbf{X}_n\} \to c(x)$ in probability or w.p.1 to prove the weak or strong consistency of $\tilde{H}_{n,m}$. Now let Y have the cdf F_n given $\mathbf{X}_n$ and Z be a random variable which is independent of $\mathbf{X}_n$ and Y and has the cdf K. Then it is easily seen that $Y + h_n Z$ has the same distribution as $\tilde{Y}_1$ given $\mathbf{X}_n$ (Falk and Reiss, 1989). Therefore

$$\begin{aligned} D_n &= mP\{Y > a_m(x - a_m^{-1} h_n Z) + b_m \mid \mathbf{X}_n\} \\ &= mE_Z E\{P\{Y > a_m(x - a_m^{-1} h_n Z) + b_m \mid \mathbf{X}_n\} \mid Z\} \end{aligned}$$

$$= \int_{\mathbf{R}} \tilde{D}_n(z)K(dz)$$

where E_Z denotes the expectation with respect to Z.

Suppose (i) and (ii) hold. Since the support of K is bounded above, there exists $U \in \mathbf{R}$ such that $Z(\omega) \leq U$ for each $\omega \in \Omega$. Let $\{t_i\}_i^\infty$ be a sequence of real numbers such that $t_i \uparrow U$. Since $\tilde{D}_n$ is nondecreasing in z, for each i,

$$\tilde{D}_n(t_i)K(t_i) \leq \int_{\mathbf{R}} \tilde{D}_n(z)K(dz) \leq \tilde{D}_n(U)$$

By Proposition 1, $\tilde{D}_n(U) \rightarrow c(x)$ w.p.1 and $\tilde{D}_n(t_i)K(t_i) \rightarrow c(x)K(t_i)$ w.p.1 for each i. By letting $i \rightarrow \infty$, we have $\int_{\mathbf{R}} \tilde{D}_n(z)K(dz) \rightarrow c(x)$, w.p.1. Now Lemma 2.1 yields (2.28) w.p.1.

Now suppose (i) and (ii) hold in probability. By the same method used in Step 2 in the Proof of Lemma 2.2, one can show that (2.28) holds in probability. □

2.5 Multivariate Extremes

Results in Section 2.3 can be extended to multivariate extremes. Let $\mathbf{R}^d$ be the set of d-dimensional real vectors and $\mathbf{R}^{d+}$ be the set of d-dimensional real vectors whose components are positive. Let $\mathbf{X}_i := (X_i^{(1)}, X_i^{(2)}, \cdots, X_i^{(d)})$, $i = 1, 2, \cdots, n$ be iid random vectors in $\mathbf{R}^d$ with a cdf
$F(x_1, x_2, \cdots, x_d)$ and $X_{1:n}^{(j)} \leq X_{2:n}^{(j)} \leq \cdots \leq X_{n:n}^{(j)}$ be the order statistics of jth components of $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$. Much of the literature deals with asymptotic theory of the component-wise maximum $\mathbf{M}_n := (X_{n:n}^{(1)}, X_{n:n}^{(2)}, \cdots, X_{n:n}^{(d)})$, with a notable exception of Galambos (1975) who investigated the limit distribution of $(X_{n-i_1:n}^{(1)}, X_{n-i_2:n}^{(2)}, \cdots, X_{n-i_d:n}^{(d)})$. In this section, relations and operations for vectors are taken componentwise. Define

$$H_{\mathrm{j}(k)}(x_1, x_2, \cdots, x_k) = P\{X_1^{(j_1)} > x_1, X_1^{(j_2)} > x_2, \cdots, X_1^{(j_k)} > x_k\}$$

where $\mathbf{j}(k) = (j_1, j_2, \cdots, j_k)'$. In the next theorem, G_j denotes the jth marginal cdf of G.

Theorem 2.11 (Galambos, 1975) *There exist* $\mathbf{a}_n = (a_n^{(1)}, a_n^{(2)}, \cdots, a_n^{(d)}) \in \mathbf{R}^{d+}$ *and* $\mathbf{b}_n = (b_n^{(1)}, b_n^{(2)}, \cdots, b_n^{(d)}) \in \mathbf{R}^d$ *such that*

$$P\{\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n) \leq \mathbf{x}\} \to G(\mathbf{x}) \tag{2.29}$$

for each $x \in C_G$ *where* G *is a nondegenerate cdf, if and only if for each* $1 \leq j_1 < j_2 < \cdots < j_k \leq d$ *and for each* $\mathbf{x}$ *for which* $G_j(x_j) > 0$, $1 \leq j \leq d$, *the limits*

$$\lim_{n\to\infty} nH_{\mathbf{j}(k)}(a_n^{(j_1)}x_1 + b_n^{(j_1)}, \cdots, a_n^{(j_k)}x_k + b_n^{(j_k)}) = h_{\mathbf{j}(k)}(x_1, x_2, \cdots, x_k) \tag{2.30}$$

are finite, and the function

$$\tilde{G}(\mathbf{x}) := \exp\{\sum_{k=1}^{d}(-1)^k \sum_{1\leq j_1<j_2<\cdots<j_k\leq d} h_{\mathbf{j}(k)}(x_1, x_2, \cdots, x_k)\} \tag{2.31}$$

is a nondegenerate cdf. Further, the actual limit cdf of $\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n)$ *coincides with* $\tilde{G}$.

Obtaining the value of $\tilde{G(\mathbf{x})}$ analytically without further assumptions is practically impossible, because each $h_{\mathbf{j}(k)}(x_1, x_2, \cdots, x_k)$ depends on the dependence structure of $X_1^{(j_1)}, X_1^{(j_2)}, \cdots, X_1^{(j_k)}$. But the bootstrap enables us to estimate the cdf of $\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n)$ as is shown below.

Let $F_n(\cdot) = n^{-1}\sum_{i=1}^{n} I(\mathbf{X}_i \in \cdot)$ be the empirical distribution of $\{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}$. Let $\mathbf{Y}_i$, $i = 1, 2, \cdots, m$ be iid random vectors in $\mathbf{R}^d$ with the cdf F_n and let $\mathbf{M}_m^Y = (Y_{m:m}^{(1)}, Y_{m:m}^{(2)}, \cdots, Y_{m:m}^{(d)})$. Define

$$H_{n,m}(\mathbf{x}, \omega) = P\{\mathbf{a}_m^{-1}(\mathbf{M}_m^Y - \mathbf{b}_m) \leq \mathbf{x} \mid \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}.$$

Theorem 2.12 *Suppose that (2.29) holds for some nondegenerate cdf G. If $m=o(n)$, then*

$$\sup_{x\in\mathbf{R}^d} \mid H_{n,m}(\mathbf{x},\cdot) - G(\mathbf{x}) \mid \rightarrow 0 \tag{2.32}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $\lambda \in (0,1)$ then (2.32) is true w.p.1.

Proof. Define

$$\begin{aligned} H_{n,m;\,\mathbf{j}(k)}(x_1,\cdots,x_k) &= P\{Y_1^{(j_1)} > x_1,\cdots,Y_1^{(j_k)} > x_k \mid \mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n\} \\ &= \frac{1}{n}\sum_{i=1}^{n} I(X_i^{(j_1)} > x_1,\cdots,X_i^{(j_k)} > x_k). \end{aligned}$$

Then

$$\begin{aligned} & E\{mH_{n,m;\,\mathbf{j}(k)}(a_m^{(j_1)}x_1 + b_m^{(j_1)},\cdots,a_m^{(j_k)}x_k + b_m^{(j_k)})\} \\ = \;& mH_{\mathbf{j}(k)}(a_m^{(j_1)}x_1 + b_m^{(j_1)},\cdots,a_m^{(j_k)}x_k + b_m^{(j_k)}) \\ \rightarrow \;& h_{\mathbf{j}(k)}(x_1,x_2,\cdots,x_k) \end{aligned}$$

and

$$\begin{aligned} & Var\{mH_{n,m;\,\mathbf{j}(k)}(x_1,x_2,\cdots,x_k)\} \\ \leq \;& (\frac{m}{n})^2 nH_{\mathbf{j}(k)}(a_m^{(j_1)}x_1 + b_m^{(j_1)},\cdots,a_m^{(j_k)}x_k + b_m^{(j_k)}) \\ \xrightarrow{P} \;& 0 \end{aligned}$$

by Theorem 2.11. Therefore

$$mH_{n,m;\,\mathbf{j}(k)}(x_1,x_2,\cdots,x_k) \xrightarrow{P} h_{\mathbf{j}(k)}(x_1,x_2,\cdots,x_k).$$

Thus

$$
\begin{aligned}
& H_{n,m}(\mathbf{x}) \\
= \; & F_n^m(a_m^{(j_1)}x_1 + b_m^{(j_1)}, \cdots, a_m^{(j_k)}x_k + b_m^{(j_k)}) \\
= \; & \{1 - m^{-1}\sum_{k=1}^{d}(-1)^k \sum_{1 \le j_1 < j_2 < \cdots < j_k \le d} mH_{n,m;\,\mathbf{j}(k)}(a_m^{(j_1)}x_1 + b_m^{(j_1)}, \cdots, a_m^{(j_k)}x_k + b_m^{(j_k)})\}^m \\
\xrightarrow{P} \; & \exp\{\sum_{k=1}^{d}(-1)^k \sum_{1 \le j_1 < j_2 < \cdots < j_k \le d} h_{\mathbf{j}(k)}(x_1, x_2, \cdots, x_k)\}
\end{aligned}
$$

where the second equality is due to Galambos (1975), and can be derived by the inclusion-exclusion formula. The Proof for w.p.1 convergence is similar to that of Theorem 2.5. □

It is known that if (2.29) holds for some $\mathbf{a}_n \in \mathbf{R}^{d+}$ and $\mathbf{b}_n \in \mathbf{R}^d$, then each marginal cdf of $F(x_1, x_2, \cdots, x_d)$ belongs to the domain of attraction of one of the distributions Λ, Φ_α, and Ψ_α (e.g., Galambos, 1978, pp. 294). Therefore components of $\mathbf{a}_n$ and $\mathbf{b}_n$ can be chosen as in Theorem 2.3. Thus once the domain of attraction to which each marginal belongs is known, $\hat{\mathbf{a}}_n$ and $\hat{\mathbf{b}}_n$ can be defined as in Theorem 2.7 and the bootstrap distribution with estimated normalizing constants is defined as

$$\hat{H}_{n,m}(\mathbf{x}) = P\{\hat{\mathbf{a}}_m^{-1}(\mathbf{M}_m^Y - \hat{\mathbf{b}}_m) \le \mathbf{x} \mid \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}.$$

An argument similar to the Proof of Theorem 2.6 and 2.7 yields the consistency results for $\hat{H}_{n,m}(\mathbf{x})$.

3. BOOTSTRAPPING EXTREMES OF DEPENDENT RANDOM VARIABLES

3.1 Asymptotic Theory of Extremes of Stationary Processes

The classical extreme value theory outlined in section 2.1 can be extended to a wide class of dependent sequences of random variables. This section outlines the asymptotic theory of extremes of stationary processes by following Chapter 3 of Leadbetter *et al.* (1983).

Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process, i.e., for each $k \in \mathbf{N}$, the process $\{X_i\}_{i=k}^{\infty}$ has the same distribution as $\{X_i\}_{i=1}^{\infty}$. For some properties of i.i.d. random variables to hold for stationary processes, we often need the "weak dependence condition" that demands the dependence between X_i and X_j to decay in some specified way as $| i-j |$ increases. We define some commonly used weak dependence conditions below.

Definition 3.1 A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to be *m-dependent* if X_i and X_j are independent whenever $| i-j |> m$.

Example 3.1 A m-th order moving average process $\{X_i\}_{i=1}^{\infty}$ given by

$$X_i = \alpha_0\varepsilon_i + \alpha_1\varepsilon_{i-1} + \cdots + \alpha_m\varepsilon_{i-m},$$

where $\{\varepsilon_i\}$ are i.i.d. random variables, is m-dependent.

Definition 3.2 A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to be *strong mixing* if

$$\alpha(n) := \sup\{\mid P(A \cap B) - P(A)P(B) \mid : A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+n}^{\infty}, k \geq 1\} \to 0,$$

as $n \to \infty$, where $\mathcal{F}_i^j$ =the σ-algebra generated by $X_i, \cdots, X_j, 1 \leq i \leq j \leq \infty$. $\alpha(\cdot)$ is called the strong mixing coefficient of the process.

Example 3.2 Athreya and Pantula (1986) showed that ARMA(p, q) process $\{X_i\}_{i=1}^{\infty}$ given by

$$X_i = \alpha_1 X_{i-1} + \alpha_2 X_{i-2} + \cdots + \alpha_p X_{i-p} + \beta_0 \varepsilon_i + \beta_1 \varepsilon_{i-1} + \cdots + \beta_q \varepsilon_{i-q}$$

is strong mixing if (i)$E[\{\log \mid \varepsilon_1 \mid\}^+] < \infty$, (ii) the distribution of ε_1 has a nontrivial absolutely continuous component, (iii) $(X_0, X_{-1}, \cdots, X_{1-p})$ is independent of $\{\varepsilon_i\}$, (iv) $\{\varepsilon_i\}$ are i.i.d. random variables, and (v) the roots of the characteristic equation $z^p - \alpha_1 z^{p-1} - \cdots - \alpha_p = 0$, are less than one in modulus.

If the ARMA(p, q) process satisfies (iii),(iv),(v) and the cdf of ε_1 is absolutely continuous, then it is geometrically strong mixing, i.e.,

$$\alpha(l) \leq C\rho^l$$

for some $C > 0$ and $\rho \in (0, 1)$ (Doukhan, 1994, pp. 99).

Definition 3.3 For each l with $1 \leq l \leq n - 1$, define

$$\alpha_{n,l}(u_n) = \sup\{\mid P\{X_j \leq u_n, j \in A \cup B\} - P\{X_j \leq u_n, j \in A\}P\{X_j \leq u_n, j \in B\} \mid :$$

$$A \subset \{1, \cdots, k\}, B \subset \{k+l, \cdots, n\}, 1 \leq k \leq n-l\}.$$

A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to satisfy the *condition* $D(u_n)$ if $\alpha_{n,l_n}(u_n) \to 0$ for some $l_n = o(n)$.

Example 3.3 Chernick (1981) showed that a stationary Markov chain satisfies $D(u_n)$ for any sequence $\{u_n\}$ for which $\lim_{n\to\infty} F(u_n) = 1$.

It is clear that $\alpha_{n,l}(u_n)$ is nonincreasing in l. The next lemma is useful.

Lemma 3.1 (Leadbetter *et al.*, **1983)** *Let $\alpha_{n,l}(u_n)$ be as in Definition 3.3. Then the condition $\alpha_{n,l_n}(u_n) \to 0$ for some $l_n = o(n)$ is equivalent to $\alpha_{n,[n\lambda]}(u_n) \to 0$ for each $\lambda \in (0, 1)$.*

Definition 3.4 Let $\mathbf{u}_n = (u_n^{(1)}, \cdots u_n^{(r)})'$. For each l with $1 \leq l \leq n-1$, define

$$\alpha_{n,l}(\mathbf{u}_n) = \sup\{\mid P\{X_j \leq v_j, j \in A \cup B\} - P\{X_j \leq v_j, j \in A\}P\{X_j \leq v_j, j \in B\} \mid:$$

$$A \subset \{1, \cdots, k\}, B \subset \{k+l, \cdots, n\}, 1 \leq k \leq n-l\}.$$

where each v_j is any choice of the r values $u_n^{(1)}, \cdots, u_n^{(r)}$. A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to satisfy the *condition* $D(\mathbf{u}_n)$ if, $\alpha_{n,l_n}(\mathbf{u}_n) \to 0$ as $\to \infty$ for some sequence $l_n = o(n)$.

Definition 3.5 For each n, i, j with $1 \leq i \leq j \leq n$ and a sequence $\{u_n\}$, define $\mathcal{F}_i^j(u_n)$ to be the σ-algebra generated by the events $\{X_s \leq u_n\}, i \leq s \leq j$. Also for each n and $1 \leq l \leq n-l$, write

$$\tilde{\alpha}_{n,l} = \sup\{\mid P(A \cap B) - P(A)P(B) \mid: A \in \mathcal{F}_1^k(u_n), B \in \mathcal{F}_{k+l}^n(u_n), 1 \leq k \leq n-l\}.$$

$\{X_i\}_{i=1}^{\infty}$ is said to satisfy the *condition* $\Delta(u_n)$ if $\tilde{\alpha}_{n,l_n}(u_n) \to 0$ for some $l_n = o(n)$ as $n \to \infty$.

Remark 3.1 The strong mixing condition was originally introduced by Rosenblatt (1956). The condition $D(u_n)$ was introduced by Leadbetter (1974) to derive asymptotic results of extremes for stationary processes. The condition $\Delta(u_n)$ was introduced by Hsing, Hüşler and Leadbetter (1988) to prove the convergence of the exceedance point process. Note that $D(u_n)$ is equivalent to

$$\begin{aligned} | \quad & P(\max\{X_j : j \in A \cup B\} \leq u_n) \\ - \quad & P(\max\{X_j : j \in A\} \leq u_n) P(\max\{X_j : j \in B\} \leq u_n) | \to 0 \end{aligned}$$

for each $A \subset \{1, \cdots, k\}$ and $B \subset \{k + l, \cdots, n\}$ where $1 \leq k \leq n - l$. Clearly, $m-$dependence $\Rightarrow$ strong mixing $\Rightarrow \Delta(u_n) \Rightarrow D(u_n)$.

Leadbetter (1974) also introduced the next condition.

Definition 3.6 A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to satisfy the *condition* $D'(u_n)$ if

$$\overline{\lim_{k\to\infty}}\ \overline{\lim_{n\to\infty}}\, n \sum_{j=2}^{[n/k]} P\{X_1 > u_n, X_j > u_n\} = 0.$$

Intuitively, the condition $D'(u_n)$ limits the possibility of clustering of high level exceedances of $\{X_i\}_{i=1}^{\infty}$. The Extremal Types Theorem for stationary processes holds when $D(u_n)$ are satisfied for each sequence $u_n = a_n x + b_n$, $x \in \mathbf{R}$.

Theorem 3.1 (Leadbetter, 1974) *Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process and suppose that there exists $a_n > 0$ and $b_n \in \mathbf{R}$ such that*

$$P\{a_n^{-1}(X_{n:n} - b_n) \leq x\} \rightarrow G(x) \tag{3.1}$$

for each $x \in C_G$ for some nondegenerate cdf G. Suppose further that $D(u_n)$ is satisfied for $u_n = u_n(x) = a_n x + b_n$ for each $x \in \mathbf{R}$. then $G(x)$ is of the type of one of the three Gnedenko's classes listed in Theorem 2.1.

Under $D(u_n)$ and $D'(u_n)$, the asymptotic theory of extremes is considerably simplified as the following theorem shows.

Theorem 3.2 (Leadbetter, 1974) *Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process with a marginal cdf F and $\{u_n\}$ be a sequence of constants such that $D(u_n)$ and $D'(u_n)$ hold. Let $0 \leq \tau < \infty$. Then*

$$P\{X_{n:n} \leq u_n\} \rightarrow e^{-\tau} \tag{3.2}$$

iff

$$n\{1 - F(u_n)\} \rightarrow \tau. \tag{3.3}$$

Theorem 2 implies that, basically, if $D(u_n)$ and $D'(u_n)$ are satisfied for each $u_n = a_n x + b_n$, $-\infty < x < \infty$ where a_n and b_n are as in (3.1), the limit distribution is determined only by marginal cdf F and is not affected by the joint distribution of $\{X_i\}_{i=1}^{\infty}$ at all. This is indeed the case when $\{X_i\}_{i=1}^{\infty}$ are i.i.d.

Example 3.3 Suppose that $\{X_i\}_{i=1}^{\infty}$ is a stationary Gaussian sequence with zero means, unit variances, and autocovariances $\gamma(n) = EX_i X_{i+n}$. Leadbetter (1974) showed that if either $\gamma(n) \log n \rightarrow 0$ or $\sum_{n=1}^{\infty} \gamma(n)^2 < \infty$, then $D(u_n)$ and $D'(u_n)$ hold for $\{u_n\}$ for which $n(1 - F(u_n)) \to \tau$ for some $\tau > 0$.

The asymptotic theory of extremes without assuming $D'(u_n)$ has been developed by several authors. Chernick (1981) has shown that if (3.3) and $D(u_n)$ are satisfied with $u_n = u_n(\tau)$ for each $\tau > 0$, and if $P\{X_{n:n} \le u_n(\tau)\}$ has the limit for each $\tau > 0$, then

$$P\{X_{n:n} \le u_n(\tau)\} \to e^{-\theta\tau} \tag{3.4}$$

for each $\tau > 0$, for some θ with $0 \le \theta \le 1$. Because of the importance of this result, θ has a name.

Definition 3.7 (Leadbetter *et al.*, 1983) A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to have the *extremal index* θ ($0 \le \theta \le 1$) if (3.3) and (3.4) hold for each $\tau > 0$.

Example 2.2 (Hsing, 1984) Let $\{Z_i\}_{i=1}^{\infty}$ be i.i.d. rv's whose cdf is uniform. Let $X_i = \max(Z_i, Z_{i+1})$. Then $\{X_i\}_{i=1}^{\infty}$ satisfies $D(u_n)$ but does not satisfy $D'(u_n)$, and has the extremal index $1/2$.

Characterizations of θ and the asymptotic theory for extremes of the stationary process which has the extremal index θ are developed by, among others, Leadbetter (1983), Hsing (1984), Hsing, Hüsler and Leadbetter (1988), Leadbetter and Nandagopalan (1988) and Chernick, Hsing and McCormick (1991).

3.2 Efron's Bootstrap under $D(u_n)$ and $D'(u_n)$

3.2.1 Known normalizing constants

Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process with a marginal cdf F. Let $F_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \le x)$ be the empirical distribution function (edf) of the sample $\mathbf{X}_n := (X_1, X_2, \cdots, X_n)$. Let $m = m(n) \in \mathbf{N}$ be such that $m(n) \to \infty$ as $n \to \infty$. Given $\mathbf{X}_n$, let $Y_1, Y_2, \cdots, Y_m$ be conditionally i.i.d. random variables (rv's) with the cdf F_n. Now, define

$$\begin{aligned} G_n(x) &= P\{a_n^{-1}(X_{n:n} - b_n) \le x\}, && (3.5)\\ H_{n,m}(x) &= P\{a_m^{-1}(Y_{m:m} - b_m) \le x|\mathbf{X}_n\}. && (3.6)\end{aligned}$$

$H_{n,m}$ is called the Efron's bootstrap (EB) distribution of $a_n^{-1}(X_{n:n} - b_n)$. Here n and m are called the original sample size and the resample size respectively. Note that the EB ignores the dependence of the process and its resampling scheme is i.i.d. sampling.

For extremes of i.i.d. rv's, as was shown in Section 2.2, the EB distribution $H_{n,m}$ has a random limit if $m_n = n$ and thus the EB fails to provide a valid approximation to G_n. For a stationary process, a similar result holds.

Theorem 3.3 *Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process such that (3.1) holds for some $a_n > 0$, $b_n \in \mathbf{R}$ and a nondegenerate cdf G. Let $\mathbf{u}_n = (u_n^{(1)}, u_n^{(2)}, \cdots, u_n^{(r)})'$. Suppose that $D(\mathbf{u}_n)$ holds and $D'(u_n)$ holds for each sequence $u_n^{(i)} = a_n x^{(i)} + b_n$, $-\infty < x^{(i)} < \infty$. If $m = n$, then*

$$H_{n,m}(x) \xrightarrow{d} \exp\{-V(x)\}, \tag{3.7}$$

where $V(x)$ is a Poisson random variable with the mean $-\log G(x)$.

Proof. Let $\tau_{n,m} = \sum_{i=1}^{n} \varepsilon_{a_n^{-1}(X_i - b_n)}$. Under the assumptions, it is known (Leadbetter *et al.*, 1983, Theorem 5.7.2) that the point process $\tau_{n,m}$ converges weakly to $PRM(\mu)$ where μ is a measure detremined by $\mu(x, \infty] = c(x)$, $x \in \mathbf{R}$. The result follows from a variant of the continuous mapping theorem (Billingsley, 1968, Theorem 5.5) .□

It is known that in general the EB fails drastically for dependent random variables (cf. Remark 2.1 of Singh, 1981). Since i.i.d. resampling scheme of the EB fails to capture the dependence structures of the process, the EB fails to approximate the sampling distribution of statistics of interest if its limit distribution depends on the joint distribution of the process. As was noted after Theorem 3.2, the limit distribution G is completely determined by the marginal distribution F if the conditions $D(u_n)$ and $D'(u_n)$ are fulfilled. Thus it is natural to conjecture that the EB is still consistent in this situation if $m = o(n)$. In fact this conjecture is correct under $D(u_n)$, $D'(u_n)$ and an additional condition $D^-(u_n)$ defined below.

Definition 3.8 For each l with $1 \leq l$, define

$$\alpha_l^-(u_n) = \sup\{| \; P\{X_j > u_n, j \in A \cup B\} - P\{X_j > u_n, j \in A\}P\{X_j > u_n, j \in B\} \; |:$$

$$A \subset \{1, \cdots, k\}, B \subset \{k + l, \cdots\}, 1 \leq k, u \in \mathbf{R}\}.$$

A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to satisfy the *condition* $D^-(u_n)$ if $\alpha_{l_n}^-(u_n) \to 0$ for some $l_n = o(n)$.

Remark 3.2 Unlike the condition $D(u_n)$, the set B in Definition 3.8 run through infinity.

Remark 3.3 It is easily seen that Lemma 3.1 still holds when $\alpha_{n,l}(u_n)$ is replaced by $\alpha^-_{n,l}(u_n)$.

Theorem 3.4 *Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process such that (3.1) holds for some $a_n > 0$, $b_n \in \mathbf{R}$ and a nondegenerate cdf G. Suppose that $D(u_n)$, $D'(u_n)$ and $D^-(u_n)$ hold for each sequence $u_n(x) = a_n x + b_n$, $x \in \mathbf{R}$ and that the mixing coefficient $\alpha_l^-(u_n(x))$ of $D^-(u_n)$ satisfies $\overline{\lim}_{n\to\infty} n^2\alpha^-_{[n\lambda]}(u_n(x)) < \infty$ for each $x \in \mathbf{R}$ and $\lambda > 0$. If $m = o(n)$ and $\overline{\lim}_{p\to 0}\ \overline{\lim}_{n\to\infty}(\frac{m}{n})^2 \sum_{j=[np]+1}^{n} (n-j)r_n(j) = 0$, where*

$$\begin{aligned} r_n(j) &= Cov(I(X_1 > u_m), I(X_{j+1} > u_m)) \\ &= P(X_1 > u_m, X_{j+1} > u_m) - P^2(X_1 > u_m), \end{aligned}$$

then

$$\sup_{x\in\mathbf{R}} \mid H_{n,m}(x) - G(x) \mid \xrightarrow{P} 0. \tag{3.8}$$

Proof. Let $c(x) = -\log G(x)$. Since $H_{n,m}(x) = \{1 - \frac{m(1-F_n(u_m))}{m}\}^m$, it is enough to show that $m(1-F_n(u_m)) \xrightarrow{P} c(x)$. We have $Em(1-F_n(u_m)) = m(1-F(u_m)) \to c(x)$ and

$$\begin{aligned} Var(m(1-F_n(u_m))) &= Var(\frac{m}{n}\sum_{i=1}^{n} I(X_i > u_m)) \\ &= (\frac{m}{n})^2\{nr_n(0) + 2\sum_{j=1}^{n-1}(n-j)r_n(j)\} \\ &= A_{n,1} + A_{n,2} \text{ (say).} \end{aligned}$$

Then

$$A_{n,1} = \frac{m^2}{n}P(X_1 > u_m)\{1 - P(X_1 > u_m)\} \sim \frac{m}{n}c(x) \to 0.$$

For each $0 < p < 1$,

$$\begin{aligned}
\frac{1}{2}A_{n,2} &= (\frac{m}{n})^2 \sum_{j=1}^{[np]}(n-j)r_n(j) + (\frac{m}{n})^2 \sum_{[np]+1}^{n-1}(n-j)r_n(j) \\
&\leq \frac{m^2}{n} \sum_{j=2}^{[np]+1} P(X_1 > u_m, X_j > u_m) - \frac{m^2}{n}[np]P(X_1 > u_m)^2 \\
&\quad + (\frac{m}{n})^2 \sum_{[np]+1}^{n-1}(n-j)r_n(j) \\
&= B_{n,1} + B_{n,2} + (\frac{m}{n})^2 \sum_{[np]+1}^{n-1}(n-j)r_n(j) \text{ (say)},
\end{aligned}$$

where

$$\begin{aligned}
B_{n,1} &= (\frac{m}{n})m \sum_{j=2}^{[mp]} P(X_1 > u_m, X_j > u_m) + \frac{m^2}{n} \sum_{j=[mp]+1}^{[np]+1} P(X_1 > u_m, X_j > u_m) \\
&\leq (\frac{m}{n})m \sum_{j=2}^{[mp]} P(X_1 > u_m, X_j > u_m) + \frac{m^2}{n} \sum_{j=[mp]+1}^{[np]+1} P^2(X_1 > u_m) \\
&\quad + \frac{m^2}{n}([np] - [mp] + 1)\alpha^-_{[mp]}(u_m(x)) \\
&\sim 0 + pm^2P^2(X_1 > u_m) + pm^2\alpha^-_{[mp]}(u_m(x)). \text{ (by } D'(u_n))
\end{aligned}$$

Thus $\overline{\lim}_{n\to\infty} B_{n,1} \leq pc^2(x) + pK$ for some $K > 0$ by assumption. And $B_{n,2} \to pc^2(x)$ as $n \to \infty$. Therefore

$$\begin{aligned}
\overline{\lim}_{n\to\infty} Var\{m(1 - F_n(u_m))\} &= \overline{\lim}_{p\to 0} \{m(1 - F_n(u_m))\} \\
&\leq \overline{\lim}_{p\to 0}\{2pc^2(x) + pK\} + \overline{\lim}_{p\to 0}\, \overline{\lim}_{n\to\infty} (\frac{m}{n})^2 \sum_{[np]+1}^{n-1}(n-j)r_n(j) \\
&= 0.
\end{aligned}$$

Hence the theorem is proved.□

A simple choice of the reample size m is given by the next corollary.

Corollary 3.1 *Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process such that (3.1) holds for some $a_n > 0$, $b_n \in \mathbf{R}$ and a nondegenerate cdf G. Suppose that $D(u_n)$, $D'(u_n)$ and $D^-(u_n)$ hold for each sequence $u_n = u_n(x) = a_n x + b_n$, $x \in \mathbf{R}$ and that the mixing coefficient $\alpha_l^-(u_n(x))$ of $D^-(u_n)$ satisfy $\alpha_l^-(u_n(x)) \leq l^{-\eta}$ for some $\eta \geq 2$ for each n, l and $x \in \mathbf{R}$. If $m = o(n)$, (3.8) holds.*

Proof. By $D^-(u_n)$, $| r_n(j) | \leq \alpha_j^-(u_m(x)) = j^{-\eta}$, thus

$$\begin{aligned} \Big| (\frac{m}{n})^2 \sum_{[np]+1}^{n} (n-j) r_n(j) \Big| &\leq (\frac{m}{n})^2 \sum_{[np]+1}^{n} (n-j) j^{-\eta} \\ &= \frac{m^2}{n^\eta} \sum_{[np]+1}^{n} \frac{1}{n} (\frac{j}{n})^{-\eta} (1 - \frac{j}{n}) \end{aligned}$$

which converges to 0 since $\sum_{[np]+1}^{n} \frac{1}{n}(\frac{j}{n})^{-\eta}(1-\frac{j}{n}) \to \int_p^1 x^{-\eta}(1-x)dx < \infty$ and $\frac{m^2}{n^\eta} \leq (\frac{m}{n})^2 \to 0$ as $n \to \infty$. Therefore Theorem 3.4 yields the result.□

The strong consistency of the bootstrap can be proved for stationary processes satisfying $\Delta(u_n)$ and $D'(u_n)$. The next lemma is standard.

Lemma 3.2 *Suppose that the condition $\Delta(u_n)$ holds for a stationary process $\{X_i\}_{i=1}^{\infty}$. Let Y and Z be $\mathcal{F}_1^j(u_n)$-measurable and $\mathcal{F}_{j+l}^n(u_n)$-measurable, respectively, rv's such that $| Y | \leq M_1$ and $| Z | \leq M_2$. Then*

$$| E(YZ) - E(Y)E(Z) | \leq 4\tilde{\alpha}_{m,l} M_1 M_2.$$

Proof. See, for example, Theorem 17.2.1 of Ibragimov and Linnik (1971).□

Theorem 3.5 *Suppose that for each sequence $u_n = a_n x + b_n$, $x \in \mathbf{R}$, $\Delta(u_n)$ and $D'(u_n)$ hold for each sequence $u_n(x) = a_n x + b_n$, $x \in \mathbf{R}$, and that the mixing coefficient*

$\tilde{\alpha}_{n,l}(u_n)$ of $\Delta(u_n)$ satisfy $\tilde{\alpha}_{n,l}(u_n(x)) \leq \rho^l$ for some $\rho \in (0,1)$ for each n, l and $x \in \mathbf{R}$. If $m \sim n^\delta$ for some $\delta \in (0, \frac{2}{3})$, then

$$\sup_{x \in \mathbf{R}} \mid H_{n,m}(x) - G(x) \mid \to 0. \tag{3.9}$$

w.p.1.

Proof. Let $T_{n,m} = \sum_{i=1}^{n} I(X_i > u_m)$ and $p_m = P(X_1 > u_m)$. It suffices to show that $\frac{m}{n}(T_{n,m} - np_m) \to 0$ w.p.1. What follows is based on a blocking argument. Let r be an positive integer. Let $m \sim n^\delta$ for some $\delta \in (0, \frac{2}{3})$, and let $l_n \sim m^{1/2r+1} \sim n^{\delta/2r+1}$ and $k_n = [\frac{n}{2l_n}]$. Define $Z_j = I(X_j > u_m) - p_m$, $j = 1, 2, \cdots, n$ and define block-sums

$$\begin{aligned}
U_{i,n} &= \sum_{j=2(i-1)l_n+1}^{(2i-1)l_n} Z_j, \quad V_{i,n} = \sum_{j=(2i-1)l_n+1}^{2il_n} Z_j, \quad i = 1, 2, \cdots, k_n - 1 \\
U_{k_n,n} &= \sum_{j=2(k_n-1)l_n+1}^{(2k_n-1)l_n} Z_j, \quad V_{k_n,n} = \sum_{j=(2k_n-1)l_n+1}^{n} Z_j.
\end{aligned}$$

Then

$$\begin{aligned}
E(T_{n,m} - np_m)^{2r} &= E(\sum_{j=1}^{n} Z_j)^{2r} \\
&= E(\sum_{i=1}^{k_n} U_{i,n} + \sum_{i=1}^{k_n} V_{i,n})^{2r} \\
&\leq 2^{2r}\{E(\sum_{i=1}^{k_n} U_{i,n})^{2r} + E(\sum_{i=1}^{k_n} V_{i,n})^{2r}\}.
\end{aligned}$$

Define $A_{j,2r} = \{(x_1, x_2, \cdots, x_j) : x_i \in \mathbf{N}, x_1 + x_2 + \cdots + x_j = 2r\}$, then

$$E(\sum_{i=1}^{k_n} U_{i,n})^{2r} = \sum_{j=1}^{2r} \sum_{(\alpha_1, \cdots, \alpha_j) \in A_{j,2r}} \sum_{1 \leq i_1 < \cdots < i_j \leq k_n} EU_{i_1,n}^{\alpha_1} U_{i_2,n}^{\alpha_2} \cdots U_{i_j,n}^{\alpha_j}.$$

Since $U_{i,n}$'s are separated at least by l_n variables, it follows from Lemma 3.1 by a induction

$$\mid EU_{i_1,n}^{\alpha_1} U_{i_2,n}^{\alpha_2} \cdots U_{i_j,n}^{\alpha_j} - \prod_{s=1}^{j} EU_{i_s,n}^{\alpha_s} \mid \;\; \leq \;\; 16(j-1) l_n^{2r} \tilde{\alpha}_{n,l_n}(u_m(x))$$

$$\leq \quad 32 r l_n^{2r} \rho^{l_n}$$
$$\to \quad 0,$$

as $n \to \infty$. Since $E\{\sum_{i=1}^{l_n} I(X_i > u_m)\}^{2r} = 2r \int_0^\infty t^{2r-1} P\{\sum_{i=1}^{l_n} I(X_i > u_m) > t\}dt$ and

$$\begin{aligned} \sum_{j=1}^{l_n} j^{2r-1} P\{\sum_{i=1}^{l_n} I(X_i > u_m) > j\} &\leq \sum_{j=1}^{l_n} j^{2r-1} l_n p_m \\ &\sim l_n p_m \frac{l_n^2 r}{2r} \\ &\sim \frac{c(x)}{2r}, \end{aligned}$$

there exists $K > 0$ such that $| EU_{i_s,n}^{\alpha_s} | \leq K$ for each $s = 1, 2, \cdots, 2r$. Let $C(r) = \sum_{j=1}^{2r} \sum_{(\alpha_1,\cdots,\alpha_j)\in A_{j,2r}} \frac{K^j}{j!}$. Since $\#\{(i_1, i_2, \cdots, i_j) : 1 \leq i_2 < \cdots < i_j \leq\} = (\frac{k_n}{j}) \leq \frac{k_n^j}{j!}$,

$$\begin{aligned} E(\sum_{i=1}^{k_n} U_{i,n})^{2r} &\sim \sum_{j=1}^{2r} \sum_{(\alpha_1,\cdots,\alpha_j)\in A_{j,2r}} \sum_{1\leq i_1<\cdots<i_j\leq k_n} \prod_{s=1}^{j} EU_{i_s,n}^{\alpha_s} \\ &= \sum_{j=1}^{r} \sum_{(\alpha_1,\cdots,\alpha_j)\in A_{j,2r}} \sum_{1\leq i_1<\cdots<i_j\leq k_n} \prod_{s=1}^{j} EU_{i_s,n}^{\alpha_s} \\ &\leq \sum_{j=1}^{r} \sum_{(\alpha_1,\cdots,\alpha_j)\in A_{j,2r}} \frac{k_n^j}{j!} K^j \\ &\leq k_n^r C(r). \end{aligned}$$

Therefore

$$\begin{aligned} P\{\frac{m}{n}(T_{n,m} - np_m) > \varepsilon\} &\leq \varepsilon^{-2r} (\frac{m}{n})^{2r} E(T_{n,m} - np_m)^{2r} \\ &\leq \varepsilon^{-2r} 2^{2r+1} C(r) (\frac{m}{n})^{2r} k_n^r \end{aligned}$$

and

$$\sum_n^\infty (\frac{m}{n})^{2r} k_n^r \sim 2^{-r} \sum_n^\infty \frac{m^{2r}}{(nl_n)^r}$$

$$= \sum_{n}^{\infty} n^{-(r+\frac{\delta r}{2r+1}-2r\delta)}$$
$$< \infty$$

for sufficiently large r, since $r + \frac{\delta r}{2r+1} - 2r\delta = r(1 + \frac{\delta}{2+\frac{1}{r}} - 2\delta) > 1$ for sufficiently large r. Hence $\frac{m}{n}(T_{n,m} - np_m) \to 0$ w.p.1 by Borel-Cantelli Lemma.□

3.2.2 Unknown normalizing constants

Since a_m and b_m are unknown when F is unknown, they need to be replaced by some estimators $\hat{a}_m$ and $\hat{b}_m$. Let $\hat{H}_{n,m}(x,\omega)$ be defined as in (2.16). The same choice of $\hat{a}_m$ and $\hat{b}_m$ will work for a stationary process satisfying assumptions of Theorem 3.4 as shown in the following result.

Theorem 3.6 *Define $\hat{a}_m$ and $\hat{b}_m$ as in Theorem 2.7 according to the domain of attraction of F. Then, under the assumptions of Theorem 3.4,*

$$\sup_{x\in\mathbf{R}} \mid \hat{H}_{n,m}(x) - G(x) \mid \xrightarrow{p} 0. \tag{3.10}$$

Proof. We provide the proof only for the case $F \in D(\Psi_\alpha)$. The other cases are similar. As was shown in Theorem 2.7, it is enough to show that

$$\frac{\theta_F - \gamma_n}{\theta_F - \gamma_m} \to 0 \tag{3.11}$$

and

$$\frac{X_{n-l_n:n} - \theta_F}{\theta_F - \gamma_m} \xrightarrow{p} -1. \tag{3.12}$$

It was proved in the proof of Theorem 2.7 that (3.11) holds if $m = o(n)$. Since $m(1 - F_n(a_m x + b_m)) \xrightarrow{p} 0$ under the assumptions of Theorem 3.4, (3.12) holds by

Remark 2.3. Hence (3.10) is proved.□

The next theorem shows that the joint distribution of the finite number of normalized upper extremes can be approximated by its bootstrap distributiuon under $D(u_n)$ and $D'(u_n)$. The proof is the same as that of Theorem 2.8 because of the i.i.d. resampling scheme of the Efron's bootstrap.

Theorem 3.7 *Assume the assumptions of Theorem 3.4. Define $\hat{a}_m$ and $\hat{b}_m$ as in Theorem 2.7 according to the domain of attraction of F. Then*

$$\sup_{x_1>\cdots>x_r} \mid P\{\hat{a}_m^{-1}(Y_{m:m}-\hat{b}_m)\le x_1,\cdots,\hat{a}_m^{-1}(Y_{m-r+1:m}-\hat{b}_m)\le x_r|\mathbf{X}_n\}$$

$$-F_r(x_1,\cdots,x_r)\mid\to 0, \tag{3.13}$$

in probability. If the assumptions of Theorem 3.5 are satisfied, (3.13) is true w.p.1.

3.3 Efron's and Moving Block Bootstrap when $\{X_i\}_{i=1}^{\infty}$ has the extremal index θ

In this section asymptotic properties of the EB and the moving block bootstrap (MBB) are investigated when the process $\{X_i\}_{i=1}^{\infty}$ does not satisfy $D'(u_n)$. Instead it is assumed that $\{X_i\}_{i=1}^{\infty}$ has the extremal index θ, $0\le\theta\le 1$. In order to obtain results in this section, we introduce a weak dependence condition D^+ which is stronger than $D(u_n)$ but is weaker than the strong mixing condition.

Definition 3.9 For each $l\ge 1$, define

$$g(l)=\sup\{\ \mid\ P\{X_j\le u, j\in A\cup B\}-P\{X_j\le u,j\in A\}P\{X_j\le u,j\in B\}\mid,$$

$$| \quad P\{X_j > u, j \in A \cup B\} - P\{X_j > u, j \in A\}P\{X_j > u, j \in B\} \mid :$$
$$A \subset \{1, \cdots, k\}, B \subset \{k+l, \cdots\}, 1 \leq k, u \in \mathbf{R}\}.$$

A stationary process $\{X_i\}_{i=1}^{\infty}$ is said to satisfy the *condition* D^+ if $g(l) \to 0$ as $l \to \infty$.

3.3.1 Known normalizing constants

It is easily shown that the EB is inconsistent when $\theta \neq 1$ even if $m/n \to 0$ as follows. Assume that D^+ holds for $\{X_i\}_{i=1}^{\infty}$ with $g(l) = l^{-\eta}$, where $\eta > 1$ and that $m^2/n \to 0$. Then $| (\frac{m}{n})^2 \sum_{j=1}^{n-1}(n-j)r_n(j) | \leq \frac{m^2}{n} \sum_{j=1}^{n-1} j^{-\eta} \to 0$. Thus it follows, from the proof of Theorem 4, that $H_{n,m}(x) \xrightarrow{P} e^{-\gamma(x)}$, where $\gamma(x) = \lim_{n\to\infty} n(1 - F(a_n x + b_n))$, while $G_n(x) \to e^{-\theta\gamma(x)}$. Hence the EB is inconsistent. It is clear that the EB is inconsistent regardless of $\{m_n\}$. The reason of the failure of the EB in this case is that the extremal index θ (see Definition 3.7) is determined by the joint distribution of $\{X_i\}_{i=1}^{\infty}$ (Leadbetter, 1983; Leadbetter and Nandagopalan, 1989; and Chernick *et al.*, 1991) while the EB cannot capture any dependence structures of $\{X_i\}_{i=1}^{\infty}$. In the Efron's bootstrap sample, the dependence structure of $(X_1, X_2, \cdots, X_n)$ is completely destroyed because the Efron's bootstrap ignores the order of variables in $(X_1, X_2, \cdots, X_n)$.

The moving block bootstrap (MBB) was introduced by Künsch (1989), and independently by Liu and Singh (1992) in order to overcome the drawback of the Efron's bootstrap for dependent observations, and they showed that the MBB provides a valid approximation to the sampling distribution of the sample mean of a stationary process. In the following, it will be shown that the MBB also works for the maximum of a stationary process. Let $M_n = \max\{X_1, X_2, \cdots, X_n\}$. The MBB method for $a_n^{-1}(M_n - b_n)$ can be described as follows: Let $\{r_n\}$ and $\{k_n\}$ be sequences of integers such that $1 \le r_n \le n$. Define blocks of length $r = r_n$,

$$\mathbf{X}_{i,r} = (X_i, X_{i+1}, \cdots, X_{i+r-1}), \quad i = 1, 2, \cdots, n-r+1.$$

Next, a sample $(\mathbf{X}^*_{1,r}, \mathbf{X}^*_{2,r}, \cdots, \mathbf{X}^*_{k,r})$ (*the MBB sample*) is randomly drawn with replacement from $(\mathbf{X}_{1,r}, \mathbf{X}_{2,r}, \cdots, \mathbf{X}_{n-r+1,r})$. Thus in the MBB sample, the dependence structure is preserved at least within each block. Define the MBB maximum

$$M^*_m = \max\{\mathbf{X}^*_{1,r}, \mathbf{X}^*_{2,r}, \cdots, \mathbf{X}^*_{k,r}\} \tag{3.14}$$

where $m = m_n = k_n r_n$. Now define the *MBB distribution* of $a_n^{-1}(M_n - b_n)$ by

$$H_{n, r_n, k_n}(x) = P\{a_m^{-1}(M^*_m - b_m) \le x \mid \mathbf{X}_n\}. \tag{3.15}$$

We need the following variant of Lemma 2.1 of Leadbetter (1983). The proof is almost identical to that of the Lemma 2.1 of Leadbetter (1983), but it is included for the sake of completeness.

Lemma 3.3 *Let $\{X_i\}_{i=1}^{\infty}$ be a stationary process satisfying D^+. Let $\{u_n\}$ be a sequence of real numbers and $\{k_n\}$ be a sequence of integers such that $k_n = o(n)$. Suppose that there exists a sequence $\{l_n\}$ of integers such that $l_n \to \infty$, $n^{-1}k_n l_n \to 0$ and $k_n g(l_n) \to 0$. Then*

$$P\{M_n \le u_n\} - P^{k_n}\{M_{r_r} \le u_n\} \to 0, \tag{3.16}$$

where $r_n = [\frac{n}{k_n}]$.

Proof. We assume that $n(1 - F(u_n))$ is bounded, which is not necessary (cf. Leadbetter, 1974) but is the case when this lemma is applied in the next theorem.

Let $\{l_n\}$ and $\{k_n\}$ satisfy assumptions. Divide the integers $1, \cdots, n$ into intervals $I_1, I_1^*, I_2, I_2^*, \cdots, I_{k_n}, I_{k_n}^*$ where

$$I_1 = (1, 2, \cdots, r_n - l_n), I_1^* = (r_n - l_n + 1, \cdots, r_n),$$

$$I_2 = (r_n + 1, \cdots, 2r_n - l_n), I_2^* = (2r_n - l_n + 1 \cdots, 2r_n),$$

$$\cdots$$

$$I_{k_n} = ((k_n - 1)r_n + 1, \cdots, k_n r_n - l_n), I_{k_n}^* = (k_n r_n - l_n + 1 \cdots, n).$$

Thus each interval I_j contains $r_n - l_n$ integers and each I_j^* except $I_{k_n}^*$ contains l_n integers, and $I_{k_n}^*$ contains $n - k_n r_n + l_n \leq k_n + l_n$ (since $r_n = [n/k_n]$) integers. For a set of integers I, define $M(I) = max\{X_i, i \in I\}$. Then

$$\begin{aligned}
0 &\leq P(\cap_{j=1}^{k_n}\{M(I_j) \leq u_n\}) - P\{M_n \leq u_n\} \\
&\leq P\{\cup_{j=1}^{k_n} M(I_j^*) > u_n\} \\
&\leq (k_n - 1)P\{M(I_1^*) > u_n\} + P\{M(I_{k_n}^* > u_n\} \text{ (Stationarity)} \\
&\leq [(k_n - 1)l_n + (k_n + l_n)]P\{X_1 > u_n\} \\
&\leq K\frac{k_n(l_n + 1)}{n} \to 0 \qquad (3.17)
\end{aligned}$$

as $n \to \infty$, where $K > 0$ is a constant. It follows from $D(u_n)$ by a induction that

$$\begin{aligned}
\mid P(\cap_j^{k_n}\{M(I_j) \leq u_n) - P^{k_n}\{M(I_1) \leq u_n\} \mid &\leq k_n g(l_n) \\
&\to 0. \qquad (3.18)
\end{aligned}$$

Finally

$$\begin{aligned}
\mid P^{k_n}\{M(I_1) \leq u_n\} - P^{k_n}\{M_{r_n} \leq u_n\} \mid &\leq k_n[P\{M(I_1) \leq u_n\} - P^{k_n}\{M_{r_n} \leq u_n\}] \\
&= k_n P\{M(I_1) < u_n \leq M(I_1^*)\}
\end{aligned}$$

$$\leq k_n l_n P\{X_1 > u_n\}$$
$$\leq K\frac{k_n l_n}{n} \to 0. \tag{3.19}$$

The result follows by combining (3.17), (3.18) and (3.19). □

The next result gives conditions under which the MBB is consistent.

Theorem 3.8 *Suppose that a stationary process $\{X_i\}_{i=1}^{\infty}$ has the extremal index θ, $0 \leq \theta \leq 1$ and satisfies D^+. Let $\{r_n\}$ and $\{k_n\}$ be sequence of integers and let $m = m_n = k_n r_n$. Suppose that $\frac{k_n^2 r_n}{n} \to 0$, $\frac{k_n^2}{n}\sum_{i=1}^{n-2r_n} g(i) \to 0$ where $g(\cdot)$ is the mixing coefficient of D^+, and there exists a sequence of integers $\{l_n\}$ such that $\frac{l_n}{r_n} \to 0$ and $k_n g(l_n) \to 0$. Then*

$$\sup_{x\in\mathbf{R}} \mid H_{n,r_n,k_n}(x) - G(x) \mid \xrightarrow{P} 0. \tag{3.20}$$

Proof. Define the blockwise maximum $M(i)$ of i-th block by

$$M(i) = \max\{X_i, X_{i+1}, \cdots, X_{i+r_n-1}\},\ i = 1, 2, \cdots, n - r_n + 1.$$

Then, draw a sample $(M^*(1), M^*(2), \cdots, M^*(k_n))$ randomly with replacement from $\{M(1), M(2), \cdots, M(n - r_n + 1)\}$. Then clearly M_m^* is distributionally (conditional on $\mathbf{X}_n$) equal to $\max\{M^*(1), M^*(2), \cdots, M^*(k_n)\}$. Let $u_n = a_n x + b_n$ be such that $n\{(1 - F(u_n))\}$ has a finite limit and write $\gamma(x) := \lim_{n\to\infty} n\{(1 - F(u_n))\}$. Thus $P\{a_n^{-1}(M_n - b_n) \leq x\} \to e^{-\theta\gamma(x)}$ for each $x \in \mathbf{R}$. Let $N = n - r_n + 1$ and $F_{n,r_n}(x) = N^{-1}\sum_{i=1}^{N} I(M(i) \leq x)$ be the edf of $\{M(1), M(2), \cdots, M(N)\}$. Then

$$\begin{aligned} H_{n,r_n,k_n}(x) &= P\{M^*(i) \leq u_m, i = 1, 2, \cdots, k_n \mid \mathbf{X}_n\} \\ &= F_{n,r_n}^{k_n}(u_m) \\ &= \{1 - \frac{k_n(1 - F_{n,r_n}(u_m))}{k_n}\}^{k_n}. \end{aligned}$$

Thus it is enough to identify the limit of $k_n(1 - F_{n,r_n}(u_m))$. From Lemma 3.3, letting m_n have the role of n, it follows that

$$P\{M_{r_n} \le u_m\} \sim P^{\frac{1}{k_n}}\{M_m \le u_m\} \sim \exp\{-\theta c(x)\frac{r_n}{m}\}$$

as $n \to \infty$. Therefore

$$\begin{aligned} Ek_n(1 - F_{n,r_n}(u_m)) &= \frac{m}{r_n} P\{M(1) > u_m\} \\ &\sim \frac{m}{r_n}(1 - \exp\{-\theta\gamma(x)\frac{r_n}{m}\}) \\ &\sim \theta c(x) \end{aligned} \tag{3.21}$$

as $n \to \infty$. Now define

$$\begin{aligned} c_m(i) &:= Cov\{I(M(1) > u_m), I(M(i+1) > u_m)\} \\ &= P\{M(1) > u_m, M(i+1) > u_m\} - P\{M(1) > u_m\}P\{M(i+1) > u_m\}. \end{aligned}$$

Then

$$\begin{aligned} Var\{k_n(1 - F_{n,r_n}(u_m))\} &= (\frac{k_n}{N})^2 Var(\sum_{i=1}^{N} I(M(i) \le x)) \\ &= (\frac{{k_n}^2}{N})c_m(0) + 2(\frac{k_n}{N})^2 \sum_{i=1}^{N-1}(N-i)c_m(i) \end{aligned} \tag{3.22}$$

The first term of (3.22) is

$$\begin{aligned} (\frac{{k_n}^2}{N})c_m(0) &\le (\frac{{k_n}^2}{N})P\{M(1) > u_m\} \\ &= \frac{1}{n - r_n + 1}(\frac{m_n}{r_n})^2 P\{M(1) > u_m\} \\ &\sim \frac{m_n}{nr_n}\theta c(x) \text{ (from (3.21)).} \\ &\to 0 \end{aligned}$$

If $r_n < i + 1$, then it follows from D^+ that

$$\begin{aligned} c_m(i) &= P(\cup_{j=1}^{r_n}\{X_j > u_m\} \cap \cup_{j=i+1}^{i+r_n}\{X_j > u_m\}) - P^2(\cup_{j=1}^{r_n}\{X_j > u_m\}) \\ &\leq g(i+1-r_n). \end{aligned} \tag{3.23}$$

The second term of (3.22) is proportional to

$$\begin{aligned} (\frac{k_n}{N})^2 \sum_{i=1}^{N-1}(N-i)c_m(i) &= (\frac{k_n}{N})^2 \sum_{i=1}^{r_n-1}(N-i)c_m(i) + (\frac{k_n}{N})^2 \sum_{i=r_n}^{N-1}(N-i)c_m(i) \\ &= A_{n,1} + A_{n,2} \text{ (say).} \end{aligned}$$

Then

$$\begin{aligned} | A_{n,1} | &\leq (\frac{k_n}{N})^2\{N(r_n-1) - \sum_{i=1}^{r_n-1} i\} \\ &\sim (\frac{k_n{}^2}{N})r_n - (\frac{k_n}{N})^2\frac{r_n^2}{2} \to 0 \text{ (by assumption).} \end{aligned}$$

Also

$$\begin{aligned} | A_{n,2} | &\leq (\frac{k_n}{N})^2 \sum_{i=r_n}^{N-1}(N-i)g(i+1-r_n) \\ &= (\frac{k_n}{N})^2 \sum_{i=1}^{N-r_n}(N-r_n-i+1)g(i) \\ &= (\frac{k_n}{N})^2(N-r_n) \sum_{i=1}^{n-2r_n+1} g(i) - (\frac{k_n}{N})^2 \sum_{i=1}^{n-2r_n+1}(i-1)g(i) \to 0 \end{aligned}$$

by assumption. Therefore $Var(k_n(1 - F_{n,r_n}(u_m)) \to 0$ and thus the theorem is proved.□

The sufficient condition for (3.20) to hold given in Theorem 3.8 is rather complicated and it may not be of practical use. The following Corollary of Theorem 3.7 gives a simpler condition on r_n and k_n by assuming a power law decay of the mixing coefficient $g(l)$.

Corollary 3.2 *Suppose that a stationary process $\{X_i\}_{i=1}^{\infty}$ has the extremal index θ, $0 \leq \theta \leq 1$ and satisfies D^+ with $g(l) = l^{-\eta}$ for some $\eta > 1$. If the block size r_n and the number of blocks k_n in the MBB sample satisfy $r_n = n^{\varepsilon}$, $k_n = n^{\delta}$ for some ε, δ with $0 < \varepsilon < 1$, $0 < \delta < (\varepsilon \wedge \frac{1-\varepsilon}{2})$, then (3.20) holds.*

Proof. Given ε and δ which satisfy $0 < \varepsilon < 1$ and $0 < \delta < (\varepsilon \wedge \frac{1-\varepsilon}{2})$, let γ be such that $\frac{\delta}{\eta} < \gamma < \varepsilon$ and let $l_n = n^{\gamma}$. Then $l_n r_n^{-1} \to 0$ and $k_n g(l_n) \to 0$. It is easy to see $\dfrac{k_n^2 r_n}{n} \to 0$ and $\dfrac{k_n^2}{n} \sum_{i=1}^{n-r_n} g(i) \to 0$, and thus the result follows. □

3.3.2 Unknown normalizing constants

Define the MBB distribution of $a_n^{-1}(M_n - b_n)$ with estimated normalizing constants.

$$\hat{H}_{n,r_n,k_n}(x) = P\{\hat{a}_m^{-1}(M_m^* - \hat{b}_m) \leq x \mid \mathbf{X}_n\}, \tag{3.24}$$

where M_m^* is as in (3.13) and $\hat{a}_m$ and $\hat{b}_m$ are some estimators of a_m and b_m. The following theorem shows that the MBB resample size m has to satisfy $m^2/n \to 0$ for $\hat{H}_{n,r_n,k_n}$ to be weakly consistent.

Theorem 3.9 *Suppose that a stationary process $\{X_i\}_{i=1}^{\infty}$ has the extremal index θ, $0 \leq \theta \leq 1$ and satisfies D^+ with $g(l) = l^{-\eta}$ for some $\eta > 1$. If the block size r_n and the number of blocks k_n in the MBB sample satisfy $r_n = n^{\varepsilon}$, $k_n = n^{\delta}$ for some ε, δ with $0 < \varepsilon < \frac{1}{2}$, $0 < \delta < (\varepsilon \wedge \frac{1-2\varepsilon}{2})$, then*

$$\sup_{x \in \mathbf{R}} | \hat{H}_{n,r_n,k_n}(x) - G(x) | \xrightarrow{p} 0. \tag{3.25}$$

Proof. Let $F \in D(\Psi_\alpha)$, i.e., $F^n(a_n x + b_n) \to \Psi_\alpha(x)$ or equivalently $n(1 - F(a_n x + b_n)) \to (-x)^{\alpha} = \gamma(x)$ (say), where $a_n = \theta_F - \gamma_n$ and $b_n = \theta_F$. It is clear that, from Corollary 3.2, $\sup_{x \in \mathbf{R}} \mid H_{n,r_n,k_n}(x) - G(x) \mid \xrightarrow{p} 0$. Therefore, as was shown in the proof of Theorem 2.7, it is enough to show (3.11) and (3.12). First, since $m^2/n = (r_n k_n)^2/n = n^{2\epsilon+2\delta-1} \to 0$, $m(1-F_n(a_m x+b_m)) \xrightarrow{p} \gamma(x)$ as was shown in the begining of Subsection 3.3.1. Therefore (3.12) holds from Remark 2.3. (3.11) also holds, since $m/n \to 0$. □

3.4 Some Conclusions

The Main findings in this chapter are:

1. When $\{X_i\}_{i=1}^{\infty}$ satisfy $D(u_n)$ and has the extremal index θ, the MBB is consistent if the block length r_n and the number of blocks k_n drawn are chosen as in Corollary 3.2.

2. When $\{X_i\}_{i=1}^{\infty}$ satisfy $D(u_n)$, the EB is inconsistent regardless of the resample size $\{m_n\}$ unless $\theta = 1$ (the case $D'(u_n)$ holds).

Therefore the MBB provides a valid approximation to the cdf of the normalized maximum for a wider class of stationary processes than the EB does. Results obtained in this chapter are similar to those of Lahiri (1992) who showed that the MBB approximates the distribution of the sample mean of a stationary process with a heavy tailed marginal cdf if the MBB sample size m is the order of $o(n)$.

4. CONFIDENCE INTERVALS FOR ENDPOINTS OF A CDF VIA BOOTSTRAP

4.1 Description of The Problem

In this chapter, the results in Chapter 2 are applied to the problem of constructing confidence intervals for endpoints of a distribution. We consider the lower endpoint because the inference problem for the lower endpoint arises naturally in reliablity applications. Now we describe the problem. Let F be a distribution function (cdf) which satisfies

$$(i) \qquad \theta'_F \equiv \inf\{x : F(x) > 0\} \in \mathbf{R},$$

$$(ii) \qquad F(\theta'_F) = 0,$$

(iii) For each $x > 0$,

$$\lim_{h \downarrow 0} \frac{F(\theta'_F + xh)}{F(\theta'_F + h)} = x^\alpha, \tag{4.1}$$

for some $\alpha > 0$. Clearly, θ'_F is the lower endpoint of the support of F. A sufficient condition for (4.1) is that F is absolutely continuous near θ'_F with a density $f(\cdot)$ such that $f(x) \sim C(x - \theta'_F)^{\alpha-1}$ for some constant $C > 0$ as $x \downarrow \theta'_F$.

Example 4.1 Families of cdf's which satidfy (i), (ii) and (iii) (given below by their densities) are the following location and scale families:

(a) Weibull distributions

$$f(x) = C_1(\frac{x-\theta'_F}{\beta})^{\alpha-1}\exp\{-(\frac{x-\theta'_F}{\beta})^{\alpha}\}, \quad \alpha,\beta > 0, \theta'_F < x < \infty,$$

(b) Gamma distributions

$$f(x) = C_2(\frac{x-\theta'_F}{\beta})^{\alpha-1}\exp\{-\frac{x-\theta'_F}{\beta}\}, \quad \alpha,\beta > 0, \theta'_F < x < \infty, \text{and}$$

(c) Beta distributions

$$f(x) = C_3(\frac{x-\theta'_F}{\beta})^{\alpha-1}(1-\frac{x-\theta'_F}{\beta})^{\eta-1}, \quad \alpha,\beta,\eta > 0, \theta'_F < x < \theta'_F + \beta.$$

Examples in reliablity applications, for which these families of distributins are useful, can be found in the chapter 3 and 4 of Bain (1987). Under assumptions (i), (ii) and (iii), the problems of estimating θ'_F and constructing confidence intervals for θ'_F have been investigated by several authors including Robson and Whitlock (1964), Cooke (1979), Hall (1981), de Haan (1981), Weissman (1981, 1982), Loh (1984) and Csörgö and Mason (1988). The purpose of this chapter is to construct asymptotically correct confidence intervals for θ'_F by using bootstrap techniques. In this chapter results are derived for the lower endpoint but results for the upper endpoint are completely parallel.

4.2 Asymptotic Theory of Lower Extremes and Bootstrap

The asymptotic theory of lower extremes is needed in order to make inference on the lower endpoint of a cdf. Indeed, we can obtain results on the lower extremes from those on the upper extremes by transforming the random variables X_i into $(-X_i)$. For this reason we only state the results on lower extremes without proofs. Let

$X_1, X_2, \cdots X_n$ be i.i.d. random variables with a cdf F and $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the corresponding order statistics. Let $F^{-1}(u) := \sup\{x : F(x) \le u\}$ be the right continuous inverse of F.

Theorem 4.1 *Suppose that there exists $a_n > 0$, $b_n \in \mathbf{R}$, $n \ge 1$ such that*

$$P\{a_n^{-1}(X_{1:n} - b_n) \le x\} = 1 - \{1 - F(a_n x + b_n)\}^n \to G(x) \tag{4.2}$$

for each $x \in C_G$ as $n \to \infty$, where G is a nondegenerate cdf . Then G is of the type of one of the following three classes.

$$\begin{aligned} &(i) \quad \Lambda^*(x) = 1 - \exp(-e^x) \quad x \in \mathbf{R}, \\ &(ii) \quad \Phi^*_\alpha(x) = \begin{cases} 1 - \exp(-(-x)^{-\alpha}) & x \le 0 \\ 1 & x > 0, \end{cases} \\ &(iii) \quad \Psi^*_\alpha(x) = \begin{cases} 0 & x \le 0 \\ \exp(-x^\alpha) & x > 0, \end{cases} \end{aligned} \tag{4.3}$$

where $\alpha > 0$.

The following theorem gives necessary and sufficient conditions for the cdf F belong to the domain of attraction of each of the three types.

Theorem 4.2 (Gnedenko, 1943) *(i) $F \in D(\Lambda^*)$ iff there exists strictly positive function g such that*

$$lim_{t \downarrow \theta'_F} \frac{F(t + xg(t))}{F(t)} = e^{-x}.$$

(ii) $F \in D(\Phi^*_\alpha)$ iff $\theta'_F = -\infty$ and $\lim\limits_{t \to -\infty} \dfrac{F(tx)}{F(t)} = x^{-\alpha}$ for each $x > 0$.

(iii) $F \in D(\Psi^*_\alpha)$ iff $\theta'_F > -\infty$ and $\lim\limits_{h \downarrow 0} \dfrac{F(\theta'_F + xh)}{F(\theta'_F + h)} = x^\alpha$ for each $x > 0$.

Theorem 4.3 (Gnedenko, 1943) *Normalizing constants a_n, b_n in (4.2) may be chosen as:*

$$
\begin{aligned}
&if \;\; F \in D(\Lambda^*), \;\; a_n = \xi_n - F^{-1}(\frac{1}{en}), \;\; b_n = \xi_n; \\
&if \;\; F \in D(\Phi_\alpha^*), \;\; a_n = \xi_n, \;\; b_n = 0; \\
&if \;\; F \in D(\Psi_\alpha^*), \;\; a_n = \xi_n - \theta'_F, \;\; b_n = \theta'_F,
\end{aligned} \tag{4.4}
$$

where $\xi_n = F^{-1}(\frac{1}{n})$.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_1, X_2, \cdots,$ be a sequence of i.i.d. random variables on $(\Omega, \mathcal{F})$ with distribution a $F \in D(G)$ where $G = \Lambda^*$ or Φ_α^* or Ψ_α^*. Let $m = m(n)$, $\mathbf{X}_n$, and $Y_1, Y_2, \cdots Y_m$ be defined as in Section 2.2. Now, define

$$G_n^*(x) = P\{a_n^{-1}(X_{1:n} - b_n) \le x\}, \tag{4.5}$$

$$\hat{H}_{n,m}^*(x, \omega) = P\{\hat{a}_m^{-1}(Y_{1:m} - \hat{b}_m) \le x | \mathbf{X}_n\}, \tag{4.6}$$

where $\hat{a}_m$ and $\hat{b}_m$ be some estimators of a_m and b_m based on $X_1, X_2, \cdots, X_n$. $\hat{H}_{n,m}^*(x,\omega)$ is called the bootstrap distribution of $a_n^{-1}(X_{1:n} - b_n)$. The next theorem is the consisitency result for $\hat{H}_{n,m}^*$. Let $l_n = [\frac{n}{m}]$ and $l'_n = [\frac{n}{em}]$, where $[\cdot]$ is the integer part.

Theorem 4.4 *Define*

$$
\begin{aligned}
\text{(i) if } F \in D(\Lambda^*), \;\; \hat{a}_m &= F_n^{-1}(\frac{1}{m}) - F_n^{-1}(\frac{1}{em}) = X_{l_n:n} - X_{l'_n:n}, \\
\hat{b}_m &= F_n^{-1}(\frac{1}{m}) = X_{l_n:n} \\
\text{(ii) if } F \in D(\Phi_\alpha^*), \;\; \hat{a}_m &= F_n^{-1}(\frac{1}{m}) = X_{l_n:n}, \\
\hat{b}_m &= 0, \\
\text{(iii) if } F \in D(\Psi_\alpha^*), \;\; \hat{a}_m &= F_n^{-1}(\frac{1}{m}) - \theta'_{F_n} = X_{l_n:n} - X_{1:n}, \\
\hat{b}_m &= \theta'_{F_n} = X_{1:n}.
\end{aligned}
$$

If m=o(n), then

$$\sup_{x\in\mathbf{R}} \mid \hat{H}^*_{n,m}(x,\cdot) - G(x) \mid \to 0 \tag{4.7}$$

in probability where G is as in (4.3). Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for each* $\lambda \in (0,1)$ *then (4.7) is true w.p.1 .*

The following theorem shows that the joint distribution of $a_n^{-1}(X_{1:n} - b_n), a_n^{-1}(X_{2:n} - b_n), \cdots, a_n^{-1}(X_{r:n} - b_n)$ can be bootstrapped consistently. Let

$$F_r(x_1, \cdots, x_r) \equiv \lim_{n\to\infty} P\{a_n^{-1}(X_{1:n} - b_n) \le x_1, \cdots, a_n^{-1}(X_{r:n} - b_n) \le x_r\}.$$

Theorem 4.5 *Assume the hypotheses on F and choose* $\hat{a}_m$ *and* $\hat{b}_m$ *as in Theorem 4.4 according to the domain of attraction F belongs to. If m=o(n), then*

$$\sup_{x_1 > \cdots > x_r} \mid P\{\hat{a}_m^{-1}(Y_{1:m} - \hat{b}_m) \le x_1, \cdots, \hat{a}_m^{-1}(Y_{r:m} - \hat{b}_m) \le x_r | \mathbf{X}_n\}$$

$$- F_r(x_1, \cdots, x_r) \mid \to 0, \tag{4.8}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for every* $\lambda \in (0,1)$ *then (4.8) is true w.p.1.*

Theorem 4.6 *Assume the hypothesis on F and choose* $\hat{a}_m$ *and* $\hat{b}_m$ *as in Theorem 4.4 according to the domain of attraction F belongs to. Let* $f : \mathbf{R}^r \to \mathbf{R}^l$ *be continuous a.e. with respect to* $F_r(\cdot, \cdots, \cdot)$*. If m=o(n), then*

$$\sup_{\mathbf{y}\in\mathbf{R}^l} \mid P\{f(\hat{a}_m^{-1}(Y_{1:m} - \hat{b}_m), \hat{a}_m^{-1}(Y_{2:m} - \hat{b}_m), \cdots, \hat{a}_m^{-1}(Y_{r:m} - \hat{b}_m)) \le \mathbf{y} | \mathbf{X}_n\}$$

$$- P\{f(a_n^{-1}(X_{1:n} - b_n), a_n^{-1}(X_{2:n} - b_n) \cdots, a_n^{-1}(X_{r:n} - b_n)) \le \mathbf{y}\} \mid \to 0 \tag{4.9}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for each* $\lambda \in (0,1)$ *then (4.9) is true w.p.1.*

Next we define the smoothed bootstrap and state its consistency results. Let $\tilde{F}_n$ be the kernel estimator of F and K and h_n be its continuous kernel cdf and the bandwidth respectively, as in Section 2.4. Given $\mathbf{X}_n$, let $\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_m$ be conditionally i.i.d. random variables having the cdf $\tilde{F}_n$. Define the smoothed bootstrap distribution of $a_n^{-1}(X_{1:n} - b_n)$ by

$$\tilde{H}^*_{n,m}(x, \omega) = P\{a_m^{-1}(\tilde{Y}_{1:m} - b_m) \le x | \mathbf{X}_n\}. \tag{4.10}$$

We have consistency results of $\tilde{H}^*_{n,m}$ under some conditions on K and h_n.

Theorem 4.7 *Suppose that (4.2) holds and the support of K is bounded below. Assume:*

(*i*) *There exists a countable dense set $D \subset \mathbf{R}$ such that for each $x \in D$,*

$$mF_n(a_m x + b_m) \to c(x), \quad w.p.1,$$

(*ii*) $h_n / a_m \to 0, \quad w.p.1.$

Then

$$\sup_{x \in \mathbf{R}} | \tilde{H}^*_{n,m}(x, \cdot) - G(x) | \to 0 \tag{4.11}$$

w.p.1. If (i) and (ii) hold in probability then so does (4.11).

Remark 4.1 From Theorem 2.5, the condition $m = o(n)$ and the condition $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for every $\lambda \in (0,1)$ are sufficient for (i) to hold in probability and w.p.1 respectively.

Remark 4.2 It can be easily seen that Theorem 2.6 holds when $Y_{1:m}$ is replaced by $\tilde{Y}_{1:m}$ and therefore Theorem 4.7 still holds when a_m and b_m are replaced by $\hat{a}_m$ and

$\hat{b}_m$ for which $\hat{a}_m/a_m \to 1$ and $a_m^{-1}(\hat{b}_m - b_m) \to 0$. The smoothed bootstrap version of Theorem 4.6 also holds.

Remark 4.3 Any lower endpoint L of the support of K yields the same asymptotic result in Theorem 6. But for $F \in D(\Psi_\alpha^*)$ a reasonable choice is $L = 0$ because if $L < 0$ the bootstrap distribution $P\{(\tilde{Y}_{1:m} - X_{1:n})/(X_{l_n:n} - X_{1:n}) \leq x \mid \mathbf{X}_n\}$ will have a positive mass on the negative real line while $P\{(X_{1:n} - \theta'_F)/(\gamma_n - \theta'_F) \leq x\}$ does not.

4.3 Confidence Intervals for Endpoints of a Cdf

Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with a cdf $F \in D(\Psi_\alpha^*)$. As was explained in Weissman (1981) or Example 4.1, three parameter Weibull, three parameter gamma and three parameter beta distributions all belong to $D(\Psi_\alpha^*)$ for some $\alpha > 0$. This fact shows the wide applicability of the assumptions (i),(ii) and (iii) given in Section 4.1 to life-testing problems. In this section, we consider the problem of constructing confidence intervals for θ'_F.

For F continuous, Weissman (1981) defined

$$R_{k,n} = \frac{X_{1:n} - \theta'_F}{X_{k:n} - X_{1:n}}, \quad k \geq 2,$$

and showed that for each $x \geq 0$,

$$P(R_{k,n} \leq x) \to 1 - (1 - (\frac{x}{1+x})^\alpha)^{k-1} \equiv G_{k,\alpha}(x).$$

He proposed asymptotically correct confidence intevals for θ_F based on $R_{k,n}$ and percentiles of the limiting distribution $G_{k,\alpha}$ and showed that, for these confidence intervals, larger k value produce shorter expected length with respect to $G_{k,\alpha}$. Since $G_{k,\alpha}$ depends on α, this method can be used only when α is known or estimated from the data. For the case α is unknown, Weissman (1982) defined

$$W_{k,n} = \log \frac{X_{k:n} - \theta'_F}{X_{1:n} - \theta'_F} \Big/ \sum_{i=1}^{k-1} \log \frac{X_{k:n} - \theta'_F}{X_{i:n} - \theta'_F}, \quad k \geq 3$$

and

$$Q_{k,l,n} = \log \frac{X_{l:n} - \theta'_F}{X_{1:n} - \theta'_F} \Big/ \log \frac{X_{k:n} - \theta'_F}{X_{l:n} - \theta'_F}, \quad 1 < l < k < n.$$

He showed that for each $x \geq 0$,

$$P(W_{k,n} \leq x) \to H_k(x) := \begin{cases} 0 & 0 \leq x \leq (k-1)^{-1} \\ \sum_{j=0}^{k^*} (-1)^j \binom{k-1}{j} (1 - jx)^{k-2} & (k-1)^{-1} \leq x \leq 1, \end{cases}$$

where $k^* = [1/x]$ and for each $x \in [0,1]$,

$$P(Q_{k,l,n} \leq x) \to G_{k,l}(x) := \sum_{j=0}^{k-l-1} (-1)^j \binom{l-1}{j} B(l + jx, k - l)/B(l, k - l),$$

where $B(a,b)$ is the beta function. Clearly both H_k and $G_{k,l}$ are free of any unknown parameters. He examined asymptotic properties of confidence intervals based on $W_{k,n}$ and those based on $Q_{k,l,n}$ for various values of k and l and concluded that confidence intervals based on $W_{k,n}$ are better than those based on $Q_{k,l,n}$ as far as their asymptotic average lengths are concerned.

An important issue that Weissman did not discuss is the finite sample properties of $R_{k,n}$,$W_{k,n}$ and $Q_{k,l,n}$. Even though larger k produces asymptotically shorter expected length of ci's based on $R_{k,n}$, k must be much smaller than n in order for $G_{k,\alpha}$ to be a good approximation to the exact distribution of $R_{k,n}$ for a finite n. Similar arguments apply to both $W_{k,n}$ and $Q_{k,l,n}$. Thus, in order to implement Weissman's methods, it is essential to choose k such that the coverage probabilities of the resulting confidence intervals are close to or greater than the nominal coverage.

Loh (1984) was the first to consider bootstrapping $R_{k,n}$. He considered the case $m_n = n$ and pointed out the difficulty that the bootstrap analogue $(Y_{1:n} - X_{1:n})/(Y_{k:n} - Y_{1:n})$ of $R_{k,n}$ cannot be defined when $Y_{k:n} = Y_{1:n}$ which does occur with positive probability. One way of resolving this difficulty is defining the bootstrap analogue of $R_{k,n}$ by

$$R^*_{k,n,m} = \begin{cases} (Y_{1:m} - X_{1:n})/(Y_{k:m} - Y_{1:m}) & if \;\; Y_{k:m} > Y_{1:m} \\ 0 & if \;\; Y_{k:m} = Y_{1:m}. \end{cases} \tag{4.12}$$

The next lemma shows that the event $Y_{k:m} > Y_{1:m}$ happens with high probability for large n if $m/n \to 0$.

Lemma 4.1 *If F is continuous and $m/n \to 0$, then for each $k \geq 2$, $P(Y_{k:m} > Y_{1:m} \mid \mathbf{X}_n) \to 1$, w.p.1 .*

Proof. We have

$$\begin{aligned} P(Y_{k:m} > Y_{1:m} \mid \mathbf{X}_n) &\geq P(Y_{2:m} > Y_{1:m} \mid \mathbf{X}_n) \\ &= \sum_{j=1}^{n} P(Y_{2:m} > Y_{1:m}, Y_{1:m} = X_{j:n} \mid \mathbf{X}_n) \\ &= \sum_{j=1}^{n} \frac{m}{n}(1 - \frac{j}{n})^{m-1}. \end{aligned}$$

For $(j-1)/n \leq x \leq j/n$, $(1-j/n)^{m-1} \leq (1-x)^{m-1} \leq (1-(j-1)/n)^{m-1}$ and thus

$$\frac{m}{n}\sum_{j=1}^{n}(1-j/n)^{m-1} \leq m \sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} (1-x)^{m-1}\, dx \leq \frac{m}{n}\sum_{j=1}^{n}(1-(j-1)/n)^{m-1}.$$

But $m\sum_{j=1}^{n}\int_{(j-1)/n}^{j/n}(1-x)^{m-1}\, dx = m\int_0^1 (1-x)^{m-1}\, dx = 1$, therefore $\overline{\lim}_{n\to\infty} \frac{m}{n}\sum_{j=1}^{n}(1-j/n)^{m-1} \leq 1$. Since $m/n \to 0$,

$$\underline{\lim}_{n\to\infty} \frac{m}{n}\sum_{j=1}^{n}(1-\frac{j}{n})^{m-1} = \underline{\lim}_{n\to\infty}\{\frac{m}{n}\sum_{j=1}^{n}(1-\frac{j-1}{n})^{m-1} - \frac{m}{n}\}$$

$$\geq \underline{\lim}_{n\to\infty} m \sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} (1-x)^{m-1}\, dx$$
$$\geq 1.$$

Therefore $\lim_{n\to\infty} \frac{m}{n} \sum_{j=1}^{n} (1-\frac{j}{n})^{m-1} = 1$ and the proof is completed.□

The following is the consistency result for bootstrapping $R_{k,n}$.

Theorem 4.8 *Assume that F is continuous and $F \in D(\Psi_\alpha^*)$. If $m=o(n)$, then*

$$\sup_{x\in\mathbf{R}} \mid P(R^*_{k,n,m} \leq x|\mathbf{X}_n) - P(R_{k,n} \leq x) \mid \to 0 \tag{4.13}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $\lambda \in (0,1)$ then (4.13) is true w.p.1.

Remark 4.4 Although $R_{k,n}$ is not asymptotically pivotal, the bootstrap still can be used to approximate the distribution of $R_{k,n}$. This is one of the advantages of bootstrap methods and it allows us to consider a wider class of statistics to be used. Similarly, it can be shown that distributions of both $W_{k,n}$ and $Q_{k,l,n,}$ can be approximated by their bootstrap distributions.

Proof of Theorem 4.8. We have

$$P(R^*_{k,n,m} \leq x|\mathbf{X}_n) = P(\frac{Y_{1:m} - X_{1:n}}{Y_{k:m} - Y_{1:m}} \leq x, Y_{k:m} > Y_{1:m} \mid \mathbf{X}_n) + P(Y_{k:m} = Y_{1:m} \mid \mathbf{X}_n).$$

From Lemma 4.1, the second term converges to zero w.p.1 . On the set $\{\omega : Y_{k:m} > Y_{1:m}\}$,

$$\frac{Y_{1:m} - X_{1:n}}{Y_{k:m} - Y_{1:m}} = \frac{(Y_{1:m} - X_{1:n})/(X_{l_n:n} - X_{1:n})}{(Y_{k:m} - X_{1:n})/(X_{l_n:n} - X_{1:n}) - (Y_{1:m} - X_{1:n})/(X_{l_n:n} - X_{1:n})}$$

$$= \frac{\hat{a}_m^{-1}(Y_{1,m} - \hat{b}_m)}{\hat{a}_m^{-1}(Y_{k,m} - \hat{b}_m) - \hat{a}_m^{-1}(Y_{1,m} - \hat{b}_m)},$$

where $l_n = [\frac{n}{m}]$. Since $f(x_1, \cdots x_r) = x_1/(x_k - x_1)$ is continuous a.e. with respect to $F(\cdot, \cdots, \cdot)$, Theorem 4.6 yields the desired result. □

Another way of avoiding the difficulty Loh pointed out is to use smoothed bootstrap technique described in Section 2.4 and 4.1. Let K, h_n and $\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_m$ be as in section 2.4. Since $\tilde{Y}_i$'s are continuous random variables given $\mathbf{X}_n$,

$$\tilde{R}^*_{k,n,m} = (\tilde{Y}_{1:m} - X_{1:n})/(\tilde{Y}_{k:m} - \tilde{Y}_{1:m}) \tag{4.14}$$

is well defined w.p.1 . The result similar to Theorem 4.8 is:

Theorem 4.9 *Suppose that F is continuous, $F \in D(\Psi^*_\alpha)$ and the support of K is bounded below. If $m_n = o(n)$ and $h_n/a_m \xrightarrow{P} 0$, then*

$$\sup_{x \in \mathbf{R}} \mid P(\tilde{R}^*_{k,n,m} \le x | \mathbf{X}_n) - P(R_{k,n} \le x) \mid \to 0, \tag{4.15}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ for each $\lambda \in (0,1)$ and $h_n/a_m \to 0$ w.p.1 then (4.15) holds w.p.1 .

Remark 4.5 Since $\hat{a}_m/a_m \to 1$ where $\hat{a}_m = X_{l_n:n} - X_{1:n}$, $h_n/(X_{l_n:n} - X_{1:n}) \to 0$ is sufficient for $h_n/a_m \to 0$. Therefore the appropriate bandwidth h_n can be determined by the data, e.g., $h_n = n^{-1/2}(X_{l_n:n} - X_{1:n})$.

Remark 4.6 By using Theorem 4.9, one can make a bias correction for $X_{1:n}$ as an estimator of θ'_F as follows: Suppose that $m_n = o(n)$ and let η_n and $\hat{\eta}_{n,m}$ be medians of $P\{R_{k,n} \le x\}$ and $P\{\hat{R}^*_{k,n,m} \le x \mid \mathbf{X}_n\}$ respectively. Define

$$\hat{\theta}'_{n,m} = X_{1:n} - \hat{\eta}_{n,m}(X_{k:n} - X_{1:n}).$$

Then

$$P\{\hat{\theta}'_{n,m} \leq \theta'_F\} = P\{\frac{X_{1:n} - \theta'_F}{X_{k:n} - X_{1:n}} \leq \hat{\eta}_{n,m}\} \to \frac{1}{2},$$

since the limit distribution $G_{k,\alpha}$ of $R_{k,n}$ is continuous. (see Beran, 1984). Thus $\hat{\theta}'_{n,m}$ is asymptotically median unbiased, and is called a *median-bias corrected estimator* of θ'_F, a terminology Loh (1984) used. Performance of $\hat{\theta}'_{n,m}$ is not yet investigated.

4.4 Type II Censoring

An advantage of Weissman's method is that confidence intervals can be obtained even when $X_{k+1:n} \leq X_{k+2:n} \leq \cdots \leq X_{n:n}$ are censored. This makes his method useful especially in life-testing and reliability problems where censoring is very common. On the other hand, the bootstrap distributions in Theorem 4.7~4.9 depend on the whole sample $\mathbf{X}_n$ and therefore it cannot be used in censoring situations. Now we define a different resampling scheme when the sample is censored on the right. Assume that only $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{c_n:n}$, $1 \leq c_n \leq n$ are observed (type II censoring). Let $\mathbf{X}_{c_n,n} = (X_{1:n}, X_{2:n}, \cdots, X_{c_n:n})$. Let $F_{c_n,n}$ be the distribution which assigns probability $1/n$ on each $X_{i:n}, i = 1, 2, \cdots, c_n$, and $(n - c_n)/n$ on any point $U_n \geq X_{c_n:n}$, i.e.,

$$F_{c_n,n}(x) = \frac{1}{n}\sum_{i=1}^{c_n} I(X_{i:n} \leq x) + \frac{n - c_n}{n} I(U_n \leq x).$$

Let $Y'_1, Y'_2, \cdots, Y'_m$ be conditionally i.i.d. random variables given $\mathbf{X}_{c_n,n}$ with distribution $F_{c_n,n}$ and $Y'_{1:m} \leq Y'_{2:m} \leq \cdots \leq Y'_{m:m}$ be the corresponding order statistics. Define

$$H_{n,m,c_n}(x, \omega) = P\{a_m^{-1}(Y'_{1:m} - b_m) \leq x | \mathbf{X}_{c_n,n}\} \tag{4.16}$$

and call H_{n,m,c_n} the *censored bootstrap distribution* of $a_n^{-1}(X_{1:n} - b_n)$.

Theorem 4.10 *Assume that (4.2) holds and* $\underline{\lim}_{n\to\infty} c_n/n > 0$. *If* $m = o(n)$, *then the censored bootstrap distribution is weakly consistent. i.e.,*

$$\sup_{x\in\mathbf{R}} \mid H_{n,m,c_n}(x,\cdot) - G(x) \mid \to 0, \tag{4.17}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for each* $\lambda \in (0,1)$ *then the censored bootstrap distribution is strongly consistent, i.e., (4.17) is true w.p.1.*

Proof. We have

$$\begin{aligned}
& \mid mF_n(a_m x + b_m) - mF_{c_n,n}(a_m x + b_m) \mid \\
= \; & \frac{m}{n} \mid (\sum_{i=c_n+1}^{n} I(X_{i:n} \le a_m x + b_m) - (n - c_n) I(U_n \le a_m x + b_m) \mid \\
\le \; & 2\frac{m}{n}(n - c_n) I(X_{c_n:n} \le a_m x + b_m).
\end{aligned}$$

Let $S_n'(x) = \#\{i : 1 \le i \le n, X_i \le a_m x + b_m\}$, then $S_n'(x) \sim Binomial(n, p_m)$, where $p_m = P(X_1 \le a_m x + b_m)$. Since $\underline{\lim}_{n\to\infty} c_n/n > 0$, there exist $0 < \eta < 1$ such that $c_n/n - p_m > \eta > 0$ for sufficiently large n, therefore

$$\begin{aligned}
P(\frac{m}{n}(n - c_n) I(X_{c_n:n} \le a_m x + b_m) > \varepsilon) & \le P(X_{c_n:n} \le a_m x + b_m) \\
& \le P(\frac{S_n'(x)}{n} - p_m \ge \frac{c_n}{n} - p_m) \\
& \le P(\frac{S_n'(x)}{n} - p_m \ge \eta) \\
& \le \eta^{-4} E(\frac{S_n'(x)}{n} - p_m)^4 \\
& \le \eta^{-4} n^{-4} (16n + 6n(n-1)).
\end{aligned}$$

Therefore $\mid mF_n(a_m x + b_m) - mF_{c_n,n}(a_m x + b_m) \mid \to 0$, w.p.1, by Borel-Cantelli lemma. The desired results followed by Lemma 1. □

Theorem 4.10 shows that the censored bootstrap distribution approximates the distribution of normalized minimum as long as $\underline{\lim}_{n\to\infty} c_n/n > 0$. If $\underline{\lim}_{n\to\infty} c_n/n > 0$, then $1/m \leq c_n/n$ for sufficiently large n, and thus $l_n \sim n/m \leq c_n$. Therefore $\hat{a}_m = X_{l_n:n} - X_{1:n}$ and $\hat{b}_n = X_{1:n}$ are available from the censored data $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{c_n:n}$. Because $\hat{a}_m/a_m \to 1$ and $a_m^{-1}(\hat{b}_m - b_m) \to 0$ as shown in the proof of Theorem 2.7, Theorem 4.10 still holds when a_m and b_m are replaced by $\hat{a}_m$ and $\hat{b}_m$.

Now consider the problem of constructing confidence intervals for θ_F' when the data is type II censored. By the reason described in Section 4.3, we need to smooth the censored empirical distribution $F_{c_n,n}$ to obtain the bootstrap analogue of $R_{k,n}$. Define

$$\tilde{F}_{c_n,n}(x) = \frac{1}{n}\sum_{i=1}^{c_n} K((x - X_{i:n})/h_n) + \frac{n-c_n}{n} K((x - U_n)/h_n),$$

where $K : \mathbf{R} \to [0,1]$ is a continuous cdf. Given $\mathbf{X}_{c_n,n}$ let $\tilde{Y}_1', \tilde{Y}_2', \cdots, \tilde{Y}_m'$ be conditionally i.i.d. random variables having a cdf $\tilde{F}_{c_n,n}$ and $\tilde{Y}_{1:m}' \leq \tilde{Y}_{2:m}' \leq \cdots \leq \tilde{Y}_{m:m}'$ be the corresponding order statistics. Define

$$\tilde{H}_{n,m,c_n}(x,\omega) = P\{a_m^{-1}(\tilde{Y}_{1:m}' - b_m) \leq x | \mathbf{X}_{c_n,n}\} \tag{4.18}$$

and call $\tilde{H}_{n,m,c_n}$ the *smoothed censored bootstrap distribution* of $a_n^{-1}(X_{1:n} - b_n)$.

The next theorem is the consistency result for $\tilde{H}_{n,m,c_n}$.

Theorem 4.11 *Suppose that (4.2) holds,* $\underline{\lim}_{n\to\infty} c_n/n > 0$ *and the support of* K *is bounded below. If* $m = o(n)$ *and* $a_m^{-1}h_n \xrightarrow{P} 0$, *then the smoothed censored bootstrap distribution is weakly consistent, i.e.,*

$$\sup_{x\in\mathbf{R}} | \tilde{H}_{n,m,c_n}(x,\cdot) - G(x) | \to 0, \tag{4.19}$$

in probability. Moreover, if $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for each* $\lambda \in (0,1)$ *and* $a_m^{-1} h_n \to 0$, *w.p.1, then the smoothed censored bootstrap distribution is strongly consistent, i.e., (4.19) is true w.p.1.*

Proof. Use results of Theorem 4.10 and techniques used in the proof of Theorem 2.10. □

Now define the smoothed censored bootstrap analogue of $R_{k,n}$ by

$$\tilde{R}^*_{k,n,m,c_n} = (\tilde{Y}'_{1:m} - X_{1:n})/(\tilde{Y}'_{k:m} - \tilde{Y}'_{1:m}). \tag{4.20}$$

Since the proof of Theorem 4.5 and 4.6 are valid for the censored bootstrap distribution if $\underline{\lim}_{n\to\infty} c_n/n > 0$, results analogus to Theorem 4.9 in the type II censoring situations hold.

Theorem 4.12 *Suppose that* F *is continuous,* $F \in D(\Psi^*_\alpha)$, $\underline{\lim}_{n\to\infty} c_n/n > 0$ *and the support of* K *is bounded below. If* $m = o(n)$ *and* $h_n/a_m \xrightarrow{P} 0$, *then*

$$\sup_{x\in\mathbf{R}} \mid P(\tilde{R}^*_{k,n,m,c_n} \le x|\mathbf{X}_n) - P(R_{k,n} \le x) \mid \to 0, \tag{4.21}$$

in probability. Moreover, if $h_n/a_m \to 0$ *w.p.1 and* $\sum_{n=1}^{\infty} \lambda^{\frac{n}{m}} < \infty$ *for every* $\lambda \in (0,1)$ *then (4.21) holds w.p.1 .*

4.5 Automatic Selection of m

In order to implement the bootstrap methods described in earlier sections, it is desirable to know an optimal choice of m or a data-based selection rule for m. At this stage, finding a theoretically optimal m seems difficult because of unavailability of an asymptotic expansion of the cdf of the normalized extremes.

Datta and McCormick (1994) use the jackknife-after-bootstrap method as a selector of m. Jackknife-after-bootstrap variance estimates are originally developed by Efron (1992) to assess the accuracy of the bootstrap estimate. Following Efron (1992), Efron and Tibshirani (1993), the bootstrap estimates and jackknife-after-bootstrap are described as follows: Let $R(\mathbf{X}_n, F)$ be a random variable of interest ($R_{k,n}$ in our problem). Let $\mathcal{L}(R(\mathbf{X}_n, F))$ denotes the distribution of $R(\mathbf{X}_n, F)$, and let $\phi[\mathcal{L}(R(\mathbf{X}_n, F))]$ be some functional of this distribution (its 95th percentile in our problem). Now, set $\gamma(n, F) := \phi[\mathcal{L}(R(\mathbf{X}_n, F))]$, and define the botstrap statistic

$$\hat{\gamma}(m, \mathbf{X}_n) := \gamma(m, F_n) = \phi[\mathcal{L}(R(\mathbf{Y}_m, F_n))]$$

where $\mathbf{Y}_m = (Y_1, Y_2, \cdots, Y_m)$ is the bootstrap sample of the resample size m. Thus $\hat{\gamma}(m, \mathbf{X}_n)$ is an estimator of $\gamma(n, F)$. In practice, the value of $\hat{\gamma}(m, \mathbf{X}_n)$ must be approximated by the Monte Carlo method. Jackknife-after-bootstrap method estimates $Var(\hat{\gamma}(m, \mathbf{X}_n))$ by the following steps:

1. Let $F_{n,(i)}$ be the empirical distribution of $(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$, $i = 1, 2 \cdots, n$. Define $\hat{\gamma}_{(i)}(m) := \gamma(m, F_{n,(i)})$.

2. Define

$$\hat{Var}(\hat{\gamma}(m, \mathbf{X}_n)) = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\gamma}_{(i)}(m) - \hat{\gamma}_{(\cdot)}(m))^2$$

where $\hat{\gamma}_{(\cdot)}(m) = \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_{(i)}(m)$.

$\hat{Var}(\hat{\gamma}(m, \mathbf{X}_n))$ is simply the jackknife estimator of the variance of $Var(\hat{\gamma}(m, \mathbf{X}_n))$. Computational burden of $\hat{Var}(\hat{\gamma}(m, \mathbf{X}_n))$ seems huge at the first glance, but there is a nice way, proposed by Efron (1992), of computing $\hat{Var}(\hat{\gamma}(m, \mathbf{X}_n))$ by original bootstrap replications.

A selection rule of m proposed by Datta and McCormick (1994) is to select m

that attains the minimum of $\hat{Var}(\hat{\gamma}(m, \mathbf{X}_n))$. This method is intuitively appealing, but whether $\hat{Var}(\hat{\gamma}(m, \mathbf{X}_n))$ is consistent estimator of $Var(\hat{\gamma}(m, \mathbf{X}_n))$ is not known and is an important open problem.

4.6 Simulation Results

Theorem 4.9 shows the consistency of the bootstrap distribution of $R_{k,n}$, therefore the asymptotic correctness of confidence intervals for θ'_F based on the bootstrap distributions when $m = o(n)$. But in practice, where n is finite, it is desirable to have some guidelines on how to choose k in $R_{k,n}$ and resample size m. In the following, the confidence interval based on the limit distribution and the bootstrap distribution are called the *limit confidence interval* and the *bootstrap confidence interval* respectively.

Monte Carlo simulations were done in the noncensoring situation to see the performances of confidence intervals based on the bootstrap distribution of $R_{k,n}$ for various values of k and m and to make comparisons to those based on the limit distribution $G_{k,\alpha}$ of $R_{k,n}$. To construct limit confidence intervals based on $G_{k,\alpha}$, α is estimated by de Haan's (1981) estimator

$$\hat{\alpha}_{n,d} = \log d / \log\{(X_{d:n} - X_{3:n})/(X_{3:n} - X_{2:n})\}$$

where $d = d(n)$ is an integer satisfying $d \to \infty$ and $d/n \to 0$ as $n \to \infty$. In the simulation, values of 15, 20, 25 and 30 were used for d.

For bootstrap confidence intervals, the smoothed bootstrap analogue $\tilde{R}^*_{k,n,m}$ of $R_{k,n}$ given in (4.14) was used. When one tries to construct confidence intervals for θ'_F based on the bootstrap distribution of $R_{k;n}$, some consideration must be given to selection of values for k and m. Even though larger k produces confidence intervals with asymptotically shorter expected length (Weissman, 1981), k must be small enough relative to n for the limit distribution of $R_{k,n}$ to be a good approximation of the exact distribution of $R_{k,n}$. In the light of Theorem 4.9, m must be of smaller order than n for the asymptotic coverage of the bootstrap confidence interval to be equal to the nominal coverage. On the other hand, for the bootstrap distribution to give a good approximation, m must be large enough compared to k since the conditional distribution of $a_m^{-1}(Y_{k:m} - b_m)$ given $\mathbf{X}_n$ should mimic the distribution of $a_n^{-1}(X_{k:n} - b_n)$. So one can expect that for each fixed k the actual coverage of a confidence interval is close to its nominal coverage for moderate sizes of m.

4.6.1 Simulation 1

The gamma distribution with the endpoint $\theta'_F = 0$, the scale parameter 1 and the shape parameter $\alpha = 1$ (the standard exponential distribution) was used as the underlying distribution F. Sample sizes of $n = 25, 50$ and 100 were chosen. For each sample size, 1000 samples were generated by the FORTRAN NAG library and confidence intervals of nominal coverage 0.95 based on the limit distributions and the bootstrap distributions with various values of k, d, and m were obtained from each sample to compute the actual coverage (CVG) and the average length of intervals (ALT). 600 bootstrap samples were generated to compute one bootstrap distribution.

$\tilde{F}_n$ as in (2.26) was used to generate $\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_m$. The standard normal distribution truncated at 0 was used as the kernel K and the minimal spacing of $X_1, X_2, \cdots, X_n$ divided by 4 was used for the bandwidth h_n, i.e.,

$$K(x) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{1}{2}u^2} du , \quad x \geq 0,$$

$$h_n = \frac{1}{4} \min_{1 \le i \le n-1} (X_{i+1:n} - X_{i:n}).$$

Table 4.1 presents coverages and average lengths of confidence intervals based on the limit cdf $G_{k,\alpha}$ of $R_{k,n}$ when $n = 25$. Exept for $k = 2$ and $d = 15$ or $d = 20$, coverages are not close to the nominal coverage .95. It is observed that coverages decrease as k increases. This is because large values of k, such as 8 or 10, are no longer small relative to $n = 25$ and thus $G_{k,\alpha}$ does not provide a good approximation to the cdf of $R_{k,n}$.

Table 4.2 presents coverarges and average lengths of bootstrap confidence intervals for various k and m. It is observed that coverages decreases as k increases because of the reason explained for Table 4.1. For $k = 2$, coverages are greater than .95 for the most of m. When $k = 4$, coverages are smal than .95 but are close enough to .95 for $k + 4 \le m \le k + 8$ $(8 \le m \le 12)$. For $k = 4$, average lengths of intervals are considerably shorter than those for $k = 2$, which is basically because $R_{2,n}$ is more spread out than $R_{4,n}$. Therefore, for this underlying distribution and sample size, $k = 4$ and $8 \le m \le 12$ seem to be optimal choices for bootstrap confidence intervals, and they are superior to the limit confidence intervals based on $R_{2,n}$ in terms of the average length of intervals.

Table 4.3 and 4.4 presents performances of the limit and the bootstrap confidence intervals respectively, when $n = 50$. The bootstrap confidence intervals with $k = 4$ and $12 \le m \le 30$ seem optimal for $n = 50$. Coverages are greater than .95 and the average lengths are short. Results for $n = 100$ are given in Table 4.5 and 4.6. Results for the limit confidence intervals are similar to those for $n = 25, 50$. Results for the bootstrap confidence intervals are similar to those for $n = 25$.

Table 4.1: Performances of 95 % (nominal) confidence intervals based on limit laws when F is gamma (1, 1) and n=25

k		d=10	d=15	d=20
2	CVG	0.899	0.934	0.948
	ALT	2.005	1.447	1.285
4	CVG	0.829	0.879	0.897
	ALT	0.536	0.322	0.273
6	CVG	0.817	0.873	0.895
	ALT	0.387	0.240	0.202
8	CVG	0.805	0.864	0.884
	ALT	0.344	0.217	0.180
10	CVG	0.771	0.851	0.874
	ALT	0.333	0.210	0.172

Table 4.2: Performances of 95 % (nominal) bootstrap confidence intervals when F is gamma (1, 1) and n=25

k	m	k	k+2	k+4	k+6	k+8	k+10	k+12	k+14
2	CVG	0.979	0.973	0.970	0.970	0.960	0.957	0.953	0.949
	ALT	2.291	2.516	2.506	2.501	2.382	2.400	2.250	2.273
4	CVG	0.899	0.939	0.948	0.946	0.944	0.940	0.936	0.935
	ALT	0.178	0.288	0.356	0.390	0.412	0.431	0.441	0.448
6	CVG	0.849	0.911	0.927	0.932	0.931	0.933	0.933	0.928
	ALT	0.127	0.193	0.231	0.256	0.273	0.283	0.291	0.298
8	CVG	0.808	0.884	0.905	0.909	0.913	0.927	0.921	0.914
	ALT	0.112	0.163	0.191	0.209	0.222	0.230	0.242	0.248
10	CVG	0.798	0.866	0.891	0.916	0.913	0.918	0.914	0.910
	ALT	0.104	0.147	0.172	0.190	0.204	0.216	0.223	0.232

Table 4.3: Performances of 95 % (nominal) confidence intervals based on limit laws when F is gamma (1, 1) and n=50

k		d=15	d=20	d=25	d=30
2	CVG	0.921	0.938	0.943	0.948
	ALT	0.515	0.779	0.744	0.716
4	CVG	0.864	0.893	0.906	0.910
	ALT	0.086	0.167	0.156	0.148
6	CVG	0.836	0.875	0.887	0.893
	ALT	0.077	0.119	0.112	0.106
8	CVG	0.827	0.863	0.880	0.878
	ALT	0.072	0.105	0.098	0.093
10	CVG	0.823	0.853	0.871	0.875
	ALT	0.072	0.099	0.093	0.087
15	CVG	0.798	0.830	0.851	0.859
	ALT	0.074	0.094	0.087	0.081
20	CVG	0.780	0.815	0.830	0.844
	ALT	0.068	0.097	0.089	0.082

Table 4.4: Performances of 95 % (nominal) bootstrap confidence intervals when F is gamma (1, 1) and n=50

k	m	k	k+2	k+4	k+6	k+8	k+10	k+12	k+14	k+16	k+18
2	CVG	0.962	0.957	0.960	0.961	0.959	0.958	0.959	0.954	0.956	0.954
	ALT	1.117	1.789	1.963	2.138	2.147	2.141	2.227	2.241	2.215	2.147
4	CVG	0.875	0.933	0.939	0.941	0.953	0.954	0.952	0.953	0.956	0.954
	ALT	0.079	0.127	0.155	0.176	0.194	0.208	0.225	0.236	0.242	0.256
6	CVG	0.819	0.894	0.916	0.928	0.929	0.931	0.933	0.935	0.937	0.936
	ALT	0.054	0.081	0.098	0.110	0.118	0.127	0.135	0.139	0.145	0.146
8	CVG	0.789	0.860	0.887	0.897	0.901	0.920	0.913	0.917	0.922	0.928
	ALT	0.046	0.068	0.081	0.090	0.097	0.103	0.107	0.111	0.114	0.116
10	CVG	0.765	0.846	0.868	0.889	0.895	0.906	0.911	0.917	0.920	0.921
	ALT	0.042	0.061	0.073	0.081	0.087	0.092	0.095	0.098	0.100	0.102
15	CVG	0.732	0.823	0.857	0.881	0.902	0.906	0.907	0.917	0.919	0.925
	ALT	0.037	0.051	0.061	0.067	0.072	0.076	0.079	0.082	0.084	0.086
20	CVG	0.725	0.815	0.843	0.874	0.877	0.891	0.896	0.906	0.906	0.912
	ALT	0.035	0.048	0.056	0.062	0.067	0.071	0.074	0.076	0.079	0.081

Table 4.5: Performances of 95 % (nominal) confidence intervals based on limit laws when F is gamma (1, 1) and n=100

k		d=15	d=20	d=25	d=30
2	CVG	0.934	0.946	0.951	0.950
	ALT	0.411	0.375	0.363	0.355
4	CVG	0.879	0.896	0.903	0.911
	ALT	0.095	0.081	0.077	0.075
6	CVG	0.860	0.886	0.893	0.898
	ALT	0.068	0.059	0.056	0.054
8	CVG	0.861	0.887	0.894	0.898
	ALT	0.060	0.052	0.050	0.048
10	CVG	0.850	0.873	0.888	0.897
	ALT	0.057	0.049	0.047	0.045
15	CVG	0.820	0.849	0.860	0.869
	ALT	0.053	0.046	0.043	0.042
20	CVG	0.806	0.834	0.841	0.855
	ALT	0.055	0.046	0.043	0.041

Table 4.6: Performances of 95 % (nominal) bootstrap confidence intervals when F is gamma (1, 1) and n=100

k	m	k	k+2	k+4	k+6	k+8	k+10	k+12	k+14	k+16	k+18
2	CVG	0.939	0.954	0.957	0.957	0.952	0.956	0.958	0.960	0.959	0.961
	ALT	0.325	0.607	0.739	0.821	0.875	0.921	0.935	0.928	0.960	1.004
4	CVG	0.871	0.921	0.933	0.940	0.940	0.941	0.941	0.941	0.940	0.939
	ALT	0.035	0.053	0.063	0.069	0.075	0.080	0.085	0.090	0.093	0.097
6	CVG	0.824	0.898	0.919	0.925	0.938	0.939	0.932	0.930	0.931	0.934
	ALT	0.024	0.035	0.042	0.047	0.050	0.053	0.056	0.057	0.059	0.061
8	CVG	0.794	0.871	0.898	0.911	0.920	0.925	0.922	0.920	0.924	0.926
	ALT	0.020	0.030	0.035	0.039	0.042	0.044	0.046	0.048	0.049	0.050
10	CVG	0.772	0.846	0.872	0.892	0.902	0.904	0.906	0.912	0.916	0.916
	ALT	0.018	0.027	0.031	0.035	0.038	0.040	0.041	0.043	0.045	0.046
15	CVG	0.729	0.823	0.857	0.872	0.887	0.891	0.896	0.900	0.906	0.908
	ALT	0.016	0.022	0.027	0.029	0.032	0.033	0.035	0.036	0.037	0.038
20	CVG	0.708	0.804	0.834	0.856	0.870	0.883	0.887	0.894	0.902	0.906
	ALT	0.015	0.020	0.024	0.026	0.029	0.030	0.032	0.033	0.034	0.035

4.6.2 Simulation 2

Simulation 1 was designed to assess average performances of confidence intervals. However, how accurately the limit cdf and the bootstrap cdf approximate the cdf of $R_{k,n}$ in a finite sample is unclear from simulation 1. Simulation 2 was designed for this purpose and it was learned from Datta and McCormick (1994).

First, exact percentiles of the cdf of $R_{4,n}$ were computed by the Monte Carlo method with 20000 trials with the gamma distribution as an underlying distribution. Percentiles of the limit cdf $G_{k,\alpha}$ of $R_{k,n}$ were computed by assuming α is known.

Second, a sample of size $n = 200$ was generated from the gamma distribution. 5000 Bootstrap samples were drawn from the smoothed empirical distribution with the Weibull cdf as a kernel cdf K. The shape parameter α was estimated nonparametrically by de Haan's estimator $\hat{\alpha}_{n,d}$ and it was plugged into the shape parameter of the Weibull distribution. The minimal spacing divided by 4 was used as a bandwidth. A reason for this choice of K is the following: Let K be a cdf such that $K(0+) = 0$. Then, for fixed n, the local behavior of the kernel estimate $\tilde{F}_n(x) = n^{-1}\sum_{i=1}^{n} K(\frac{x - X_i}{h_n})$ near $X_{1:n}$ is the same as that of K near 0. Since our assumption is $F \in D(\Psi_\alpha^*)$, a kernel cdf K which belongs to $D(\Psi_\alpha^*)$, such as the Weibull (α, β), might perform better in a finite sample than a cdf which does not, such as the truncated normal cdf. Since α is unknown, it must be estimated and plugged into $W_\alpha(\cdot)$, the Weibull cdf. In this case, $\tilde{F}_n \in D(\Psi_{\hat{\alpha}}^*)$ for each fixed n. The choice of the Weibull cdf among the cdf's in $D(\Psi_\alpha)$ is somewhat arbitrary. It is easy to generate Weibull rv's for each $\alpha > 0$, while gamma rv's can be generated by the FORTRAN NAG library only when α is a nonnegative integer or half integer value, and this is an advantage of using Weibull rv's since generation of rv's whose cdf is K is required in resampling from $\tilde{F}_n$.

Table 4.7: Exact, Limit and Bootstrap Percentiles of cdf of $R_{4,200}$ when F is the gamma (1, 1) distribution

	Left Tail Prob.								
	.025	.05	.1	.2	.5	.8	.9	.95	.975
Exact	.00844	.01731	.03572	.07828	.25934	.70769	1.16298	1.7253	2.3597
Limit	.00840	.01695	.03451	.07168	.20630	.41520	.53584	.6316	.7076
Boot m									
10	.00808	.01480	.03276	.08048	.22936	.63339	1.08398	1.5672	2.1737
20	.00824	.01705	.03259	.07985	.26547	.68624	1.11665	1.5539	2.3913
25	.00868	.01650	.03324	.07948	.29119	.70604	1.15579	1.6584	2.6643
30	.00843	.01761	.03510	.07483	.29931	.74778	1.24294	1.8007	2.7584
40	.00856	.01737	.03556	.07648	.30018	.77607	1.33680	1.8995	3.0953
50	.00837	.01665	.03557	.07868	.32119	.85695	1.47145	2.1623	3.1063
60	.00807	.01607	.03399	.07446	.28740	.82751	1.49131	2.2393	3.3009
70	.00827	.01684	.03626	.07584	.26144	.90316	1.66983	2.5096	3.6227
80	.00755	.01594	.03440	.07347	.24911	.90507	1.70751	2.5552	3.8947
90	.00753	.01685	.03617	.07512	.23033	.83616	1.80022	2.5932	3.7681
100	.00780	.01534	.03375	.07792	.21985	.81093	1.74980	2.6543	3.7189

Table 4.7 presents the exact, limit and bootstrap percentiles for the case $\alpha = 1$, $\beta = 1$ and $n = 200$. The limit percentiles are very accurate estimates of exact percentiles when p is small, but are very poor estimates when p is large ($p > .8$). On the other hand, the bootstrap percentiles are reasonably close to exact percentiles over all $0 < p < 1$ when $m = 10 \sim 25$.

Table 4.8 presents the confidence intervals of θ'_F with the nominal coverage of .95 based on 2.5th and 97.5th percentiles obtained in Table 4.7. Inaccuracy of the limit percentiles results in a too short interval, which is a main reason for its low coverage shown in the simulation 1. Bootstrap confidence intervals are close to exact confidence intervals when $10 \leq m \leq 20$. The upper limits of confidence intervals are almost the same for each m, since both $X_{4:n} - X_{1:n}$ and 2.5th percentiles are very

Table 4.8: 95 % confidence intervals for the lower endpoint based on a sample of size $n = 200$ from the gamma (1, 1) population

	Lower limit	Upper Limit
Exact	-.06371	.00253
Limit	-.01716	.00253
Bootstrap		
m		
10	-.05847	.00254
20	-.06460	.00254
25	-.07229	.00253
30	-.07494	.00253
40	-.08443	.00253
50	-.08474	.00254
60	-.09023	.00254
70	-.09929	.00254
80	-.10695	.00256
90	-.10339	.00256
100	-.10200	.00255

small.

The same simulation was done for the gamma distribution with $\alpha = 0.5$, $\beta = 2$. Results are given in Table 4.9 and 4.10. Similar observations were made. Limit percentiles underestimate high percentiles and result in a shorter interval. Bootstrap confidence intervals are reasonably close to the exact confidence interval for $m = 10 \sim 35$.

Table 4.9: Exact, Limit and Bootstrap Percentiles of cdf of $R_{4,200}$ when F is the gamma (0.5, 2) distribution

			Left	Tail	Prob.				
	.025	.05	.1	.2	.5	.8	.9	.95	.975
Exact	.00007	.00028	.00122	.00531	.04508	.21440	.42369	.6896	1.0351
Limit	.00007	.00029	.00119	.00514	.04256	.17239	.28713	.3989	.5007
Boot									
m									
10	.00014	.00054	.00176	.00451	.02853	.14539	.30843	.5233	.8096
20	.00023	.00065	.00263	.00843	.04205	.16276	.31397	.5070	.7925
30	.00040	.00111	.00395	.01294	.06113	.20304	.40734	.6765	1.0818
35	.00037	.00109	.00372	.01255	.06721	.23681	.44356	.6943	1.0519
40	.00059	.00139	.00444	.01502	.08235	.27144	.50811	.8295	1.3106
50	.00070	.00178	.00457	.01610	.09857	.35033	.63565	1.0440	1.5938
60	.00068	.00169	.00449	.01620	.10994	.39866	.69530	1.1435	1.6284
70	.00066	.00216	.00533	.01735	.11842	.42622	.76894	1.2455	1.7993
80	.00067	.00236	.00503	.01448	.11416	.45008	.80633	1.2826	1.7544
90	.00062	.00240	.00514	.01503	.11537	.49533	.87517	1.3493	1.9036
100	.00064	.00228	.00519	.01348	.11107	.47367	.90393	1.4333	2.0819

Table 4.10: 95 % confidence intervals for the lower endpoint based on a sample of size $n = 200$ from the gamma (0.5, 2) population

	Lower limit	Upper Limit
Exact	-.00007	.00000
Limit	-.00003	.00000
Bootstrap		
m		
10	-.00006	.00000
20	-.00006	.00000
30	-.00008	.00000
35	-.00007	.00000
40	-.00009	.00000
50	-.00011	.00000
80	-.00012	.00000
100	-.00015	.00000

5. CONCLUSION AND DISCUSSION

In this thesis, it was shown that the distribution of normalized maximum can be estimated by the Efron's bootstrap for the case of (i) iid univariate rv's, (ii) iid multivariate rv's, and (iii) a stationary process satisfying the condition $D(u_n)$ and $D'(u_n)$, and by the moving block bootstrap for the case of (iv) a stationary process satisfying the condition $D(u_n)$. A sufficient condition for the consistency of these methods is $m = o(n)$ where n and m are the sample size and resample size respectively.

For the case of iid multivariate rv's, one need to specify the association structure of components of random vector $\mathbf{X}_1$ in order to obtain an approximation for the cdf of the maximum based on its limit distribution. But specifying the association is often a difficult task in practice. On the other hand, Efron's bootstrap automatically recovers the association between components of a random vector. Thus its advantage is clear in this case.

For the case of a stationary process satisfying $D(u_n)$, one needs to estimate both the extremal index θ of the sequence and α in order to obtain an approximation for the cdf of the maximum based on its limit distribution. On the other hand, moving block bootstrap is automatic in approximating the cdf of the maximum, except for the choice of the block length and the number of blocks drawn, thus it is a good alternative method to the approximation based on the limit distribution.

These two examples indicate the usefulness of resampling methods, especially for inference based on multivariate observations, and dependent obsevations, for which specifying models are much more difficult than specifying univariate cdf.

As an application, the Efron's bootstrap is used to construct confidence intervals for the lower endpoint of a cdf when observations are iid and univariate. This method is also applicable when observations are stationary and satisfy both $D(u_n)$ and $D'(u_n)$. Monte Carlo simulation results indicate that a good performance of confidence intervals based on this method depends on the choice of k and the resamle size m. Therefore, in order to implement this method, both theoretical and empirical studies of selection rules for the resample size m is particularly important. Jackknife-after-bootstrap method described in Section 4.5 might be a good candidate for a data-based selector of m. A simultaneous selection of k and m is also worth investigating.

For the implementation of the moving block bootstrap, a similar problem is the choice of the block length r_n which yields a good approximation to the cdf of the normalized maximum. To the best of our knowledge, satisfactory selection rules for r_n remains an open problem for inference problems for dependent observations. Data-based selection rules are worth investigating. It is clear that a good choice of r_n depends on the dependence structure of the process. In general, the consistency of the moving block bootstrap requires r_n to be order of $o(n)$ (e.g., Künsch, 1989), but a good choice of r_n should be large enough for blocks $(X_i, X_{i+1}, \cdots, X_{i+r_n-1})$ to preserve the dependence structure the sample $(X_1, X_2, \cdots, X_n)$ has. So, the choice of r_n adapted to some estimated dependence quantities, like autocorrelation function, might be of some importance in practice.

There are two important open problems: The problem of constructing confidence regions for the vector of lower endpoints of multivariate cdf via bootstrap is particularly important, since obtaining analytical form of the limit distribution of the multivariate version of Weissman's statistics seems very difficult. For the same reason, constructing the moving block bootstrap confidence intervals for the lower endpoint of a cdf when observations are stationary and has the extremal index $\theta \neq 1$ is important. For both problems, we conjecture that the smoothed bootstrap approximates the distributions of Weissman's statistics, but it has not been proved yet. All these problems are future research topics.

BIBLIOGRAPHY

Angus, J. (1993). Asymptotic theory for bootstrapping the extremes. *Communication in Statistics Theory and Methodology,* **22(1)**, 15-30

Arcones and Gine (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annales de l' Institut Henri Poincaré - Probabilités Statistiques,* **25** 457-481.

Athreya, K.B. (1987). Bootstrap of the mean in the infinite variance case. *Annals of Statistics,* **15**, 724-731.

Athreya, K. B. (1986). Bootstrap of the mean in the infinite variance case, II. Technical Report 86-21, Department of Statistics, Iowa State University, Ames, Iowa.

Athreya, K. B. and Fukuchi, J. (1994a). Bootstrapping extremes of i.i.d random variables. In *Proceedings of the Conference on Extreme Value Theory and Applications,* held at NIST, Maryland in 1993.

Athreya, K. B. and Fukuchi, J. (1994b). Confidence intervals for endpoints of a cdf via bootstrap. Preprint 92-11, Department of Statistics, Iowa State University, Ames, Iowa.

Athreya, K. B., Lahiri, S. N. and Wu, W. (1992) Inference for heavy tailed distributions. Preprint, Department of Statistics, Iowa State University, Ames, Iowa.

Athreya, K. B. and Pantula, S. (1986) A note on strong mixing of ARMA processes. *Statistics Probability Letters,* 4, 187-190.

Bain, L. (1978) *Statistical Analysis of Reliabilty and Life-Testing Models*, Marcell Dekker, New York.

Beran, R. (1982) Estimated sampling distributions: the bootstrap and its competitors. *Annals of Statistics*, **10**, 212-225.

Beran, R. (1984) Bootstrap methods in statisitcs. *Jber. d. Dt. Math.-Verein*, **86**, 14-30.

Bickel, P. J. (1994) Safer but still accurate bootstrap: extrapolating from lower resample sizes. The talk given in the 57th anual meeting of the institute of mathematical statistics.

Bickel, P. J. and Freedman, D. A., (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* , **9**, 1196-1217.

Billingsley, P. (1968) *Convergence of Probability Measure*, John Wiley and Sons, New York.

Bretagnolle, J. (1983) Lois limites du bootstrap de certaines functionnelles. *Annales de l' Institut Henri Poincaré - Probabilités Statistiques,* **19**, 281-296.

Chernick, M. R. (1981). On strong mixing and Leadbetter's D condition. *Journal of Applied Probability,* **18**, 764-769.

Chernick, M. R. (1981). A limit theorem for the maximum of autoregressive processes with uniform marginal distribution. *Annals of Probability,* **9**, 145-149.

Chernick, M. R., Hsing, T. and McCormick, W.P. (1991) Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability,*, **23**, 835-850.

Chow, Y. and Teicher, H. (1988). *Probability Theory, Independence, Interchangeability, Martingales.* Second Edition. Springer, Berlin.

Cooke, P. J. (1979). Statistical inference for bounds of random variables. *Biometrika,* **66**, 367-374.

Csörgö, S. and Mason, D. (1988). Simple estimators of the endpoint of a distribution. In Extreme value theory, lecture notes in statistics. **51**. 132-147. Springer-Verlag, New York.

Datta, S. (1992). On asymptotic properties of bootstrap for unstable AR(1) processes. Technical Report. Department of Statistics, University of Georgia, Althens, Georgia.

Datta, S. and McCormick, W. (1994). Bootstrap inference for a first order autoregression with positive innovations. Technical Report 94-24. Department of Statistics, University of Georgia, Althens, Georgia.

Davis, R. and Wu, W (1994). Bootstrap implementation of M-estimate for autoregression and linear regression with infinite variance. Preprint. Colorado State University, Fort Collins, Colorado.

Deheuvels, P. Mason, D. and Shorack, G. (1993). Some results on the influence of extremes on the bootstrap. *Annales de l' Institut Henri Poincaré - Probabilités Statistiques,* **29** 83-103.

Doukhan, P. (1994). Mixing: Properties and Examples. Lecture Notes in Statistics, 85. Springer-Verlag, New York.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions. *Journal of Royal Statistical Society, Series B,* **54**, 83-127.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap.* Chapman and Hall, New York.

Falk, M. and Reiss, R. D. (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Annals of Probability,* **17**, 362-371.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of American Statistical Association.***70**, 674-680.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* John Wiley and Sons, New York.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum dune seriealeatoire. *Ann. Math.* **44** 423-453.

Haan, L. de (1970). *On Regular Variation and its Application to the Weak Conver gence of Sample Extremes.* Mathematical Centre Tracts **32**, Mathematical Centre, Amsterdam.

Haan, L. de (1981). Estimation of the minimum of a function using order statistics. *Journal of American Statistical Association,* **76**, 467-469.

Hall, P. (1981). On estimating the endpoint of a distribution. *Annals of Statistics*, **9**, 556-568.

Hall, P. DiCiccio, T. and Romano, J. (1989). On smoothing and bootstrap. *Annals of Statistics*, **17**, 692-704.

Hsing. T. (1984). Point processes associated with extreme value theory. Ph. D. dissertation. Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Hsing. T., Husler, J. and Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence. *Probabilty Theory and related Fields* , **78**, 97-112.

Ibragimov and Linnik (1971). *Independent and Stationary sequences of Random Variables.* Wolters-Noordhoff, Netherlands.

Künsch, H. R., (1989). The jacknife and the bootstrap for general stationary observations, *Annals of Statistics,* **17**, 1217-124

Lahiri, S. N. (1992). On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tailed random variables. Preprint 92-3. Department of Statistics, Iowa State University, Ames, Iowa.

Leadbetter, M.R. (1974). On extreme values in stationary sequences. *Z. Wahrschein lichkeitstheorie und Verwandte Gebiete,* **28**, 289-303.

Leadbetter, M.R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete,* **65**, 291-306

Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983). *Extremes and Related Pro-*

perties of Random Sequences and Processes. Springer, Berlin.

Leadbetter, M.R. and Nandagopalan, S. (1988) On exceedance point processes for staionary sequences under mild oscillation restrictions. In Extreme value theory, lecture notes in statistics. **51**. 69-80. Springer-Verlag, New York.

Liu, R. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In Exploring the lilits of bootstrap (Eds. R. Lepage and L. Billard), John Wiley Sons, New York, 225-248.

Loh, W. (1984). Estimating an endpoint of a distribution with resampling methods. *Annals of Statistics,* **12**, 1543-1550.

Politis, D. and Romano, J. (1992). A general theory for large sample confidence regions based on subsamples under minimal assumptions, Technical Report 92-36, Department of Statistics, Purdue University, West Lafayette, Indiana.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer Verlag, New York.

Robson, D.S. and Whitlock, J. H. (1964). Estimation of a truncation point. *Biometrika,* **51**, 33-39.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Aci. USA* **42** 43-47.

Silverman, B. W. and Young, G. A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika,* **74**, 469-479.

Seneta, E. (1976). Regular Varying Functions. Lecture notes in mathematics. **508**. Springer-Verlag, New York.

Singh, K. (1981). On the asymptotic efficiency of Efron's Bootstrap. *Annals of Statistics,* **9**, 1187-1195.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works. *Communication in Statistics -Theory and Method,* **15(11)**, 3193-3203.

Weissman, I. (1981). Confidence intervals for the threshold parameter. *Communica-*

tion in Statistics -Theory and Method, **A10(6)**, 549-557.

Weissman, I. (1982). Confidence intervals for the threshold parameter II: unknown shape parameter. *Communication in Statistics - Theory and Method,* **11(21)**, 2451-2474.

ACKNOWLEDGEMENTS

Obviously, this work would not have been possible only with pulling myself up by my own "bootstrap". First of all, I wish to express my deep gratitude to my major professor Dr. Krishna B. Athreya for suggesting the problem and for advising me kindly. All through my graduate study at Iowa State, he has been a constant source of warm encouragements, inspiration and new ideas. It is his way of thinking probabilistically and of solving problems with mathematical rigor that has benefited me most.

I am also very grateful to my co-major professor Dr. Soumendra N. Lahiri. His idea is essential especially in Chapter 3 of this thesis and conversations with him opened up my interest in resampling methods for stationary processes.

I am thankful to my committee members, Dr. Yasuo Amemiya, Dr. Jay Breidt, Dr. Keneth Koehler and Dr. Ananda Weerasinghe for their constructive suggestions.

Many thanks go to Dr. Wiliam McCormick of University of Gerogia for providing me his paper with Dr. Somnath Datta. I learned many important ideas from their paper.

I am thankful to my friend, Seoung-Gon Ko, for assisting me on FORTRAN programming that produced simulation results in Chapter 4 of this thesis.

Most of all, I thank my parents, Sotaro and Akiko Fukuchi, for their support all through my graduate study both in Japan and the U.S.. I thank my lovely wife, Naoko, for supporting me in every aspect of the life in Ames. This disertation is dedicated to them with my deep appreciation.