Bootstrap and empirical likelihood methods in extremes

Yongcheng Qi

Abstract One of the major interests in extreme-value statistics is to infer the tail properties of the distribution functions in the domain of attraction of an extreme-value distribution and to predict rare events. In recent years, much effort in developing new methodologies has been made by many researchers in this area so as to diminish the impact of the bias in the estimation and achieve some asymptotic optimality in inference problems such as estimating the optimal sample fractions and constructing confidence intervals of various quantities. In particular, bootstrap and empirical likelihood methods, which have been widely used in many areas of statistics, have drawn attention. This paper reviews some novel applications of the bootstrap and the empirical likelihood techniques in extreme-value statistics.

1 Introduction

It is a great pleasure for me to write this review paper in honor of Professor Laurens de Haan on the occasion of his 70th birthday.

Dedicated to Professor Laurens de Haan on the occasion of his 70th birthday.

Y. Qi (✉)
Department of Mathematics and Statistics, University of Minnesota Duluth,
1117 University Drive, Duluth, MN 55812, USA
e-mail: yqi@d.umn.edu

Let F be a distribution function. If there exist constants $a_n > 0$ and $b_n \in R$ and a non-degenerate distribution function G such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{1.1}$$

for all continuity points x of G, then we say that F is in the domain of attraction of the distribution G and denote $F \in \mathcal{D}(G)$. It is well known that the limiting distribution G must be of the type of one of the so-called extreme-value distributions:

$$G_\gamma(x) = \exp\{-(1 + \gamma x)^{-1/\gamma}\} \text{ for } 1 + \gamma x > 0,$$

where $(1 + \gamma x)^{-1/\gamma}$ is interpreted as e^{-x} for $\gamma = 0$, and γ is called the tail index of F or the extreme-value index. For details, see, e.g., Galambos (1978), or de Haan and Ferreira (2006).

In extreme-value statistics, many important characteristics of the underlying distribution F such as the tail index, high quantiles and endpoint are related to the tail behavior of the distribution. The estimation of these characteristics is very important in extreme value statistics, and it plays a central role in predicting rare events. Based on a few very large observations, various estimators for the extreme-value index, high quantiles and the endpoint have been proposed in the literature, and their asymptotic behaviors have been investigated; see, e.g., the Hill estimator (Hill 1975) for the case $\gamma > 0$, Pickands estimator (Pickands 1975; Pereira 1993) and moment estimator (Dekkers et al. 1989) for general $\gamma \in R$, and moment-based estimators for high quantiles and the endpoint (Danielsson and Vries 1997; Ferreira et al. 2003), to just mention a few.

In this paper we review some recent developments of bootstrap techniques and empirical likelihood methods in extreme-value statistics. The rest of the paper is organized as follows. In Section 2, we introduce some applications of bootstrap techniques in estimating optimal sample fractions in estimation problems (which is largely due to the contribution of Laurens de Haan and his collaborators) and in constructing confidence intervals for the extreme value indices. In particular, we will present bootstrap procedures for the Hill estimator and the moment estimator with some technical details. In Section 3 we demonstrate how to use empirical likelihood methods to construct confidence intervals for the tail index and high quantiles of a heavy-tailed distribution.

This review paper aims to give a clear picture about the applications of the aforementioned methods in extremes. It is not the author's intention if any contributions in these areas are omitted due to limited reference resources.

2 Bootstrap Methods

The bootstrap is a data-driven method that has a very wide range of applications in statistics. Initiated by Efron (1979), the classic bootstrap approach uses Monte Carlo simulations to generate an empirical estimate for the sampling

distribution of a statistic by randomly drawing a large number of samples of the same size n from the data, where n is the size of the sample under consideration.

Although the bootstrap has been widely used in many areas, the method has its limitation in extremes. It is well known that a full-sample bootstrap does not work for the extremes. As Hall (1990) pointed out, the full-sample bootstrap statistics have a zero bias if the statistics are linear in data. Hence, in extreme-value statistics, a full-sample bootstrap approach can be invalid if the original statistics have a non-negligible bias. Angus (1993) showed that the bootstrapping distribution for the extremes does not converge to an extreme-value distribution. In fact, its limit is a random probability measure. As was illustrated by Shao and Tu (1995, Ex. 4, p. 123), a sub-sample bootstrap is still valid in this case. Further, Geluk and de Haan (2002) showed that only the sub-sample bootstrap can estimate the limiting distribution for intermediate order statistics.

Throughout let $F^-(u) = \inf\{x : F(x) \geq u\}$ be the generalized inverse function of F and U be the inverse function of $\frac{1}{1-F}$, ie., $U(x) = F^-(1 - \frac{1}{x})$. Define the $(1 - p)$-th quantile of F by $x_p = F^-(1 - p) = \inf\{x : F(x) \geq 1 - p\}$ for $0 < p < 1$ and the (right) endpoint by $\omega_F = \sup\{x : F(x) < 1\}$.

Assume that $\{X_j, j \geq 1\}$ is a sequence of independent and identically distributed (iid) random variables with distribution function $F(x)$ satisfying Eq. 1.1, and only a finite sample $\{X_j, 1 \leq j \leq n\}$ is observed. Let $X_{n,1} \leq \cdots \leq X_{n,n}$ be the order statistics of $X_1, \cdots, X_n$, and $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$ be the empirical distribution.

Assume $\{X_j^*, j \geq 1\}$ is a sequence of iid random variables with the distribution function F_n. For each integer $m \geq 1$, let $X_{m,1}^* \leq \cdots \leq X_{m,m}^*$ be the order statistics of $(X_1^*, \cdots, X_m^*)$ which is a bootstrap sample of size m from $\{X_j, 1 \leq j \leq n\}$. A full-sample bootstrap is the case when $m = n$. In contrast, a sub-sample bootstrap is the case when $m < n$ and usually the conditions $m = m(n) = o(n)$ and $m \to \infty$ are also required.

One of the common properties for the estimators of the tail index and the high quantiles is that their asymptotic biases are dependent on the fraction of the sample that is used in the estimation. The asymptotic biases increase with the increase of the sample fraction. However, when the sample fraction is small, the variances of the estimators are large. As long as accuracy of the estimation is concerned, one has to decide how to choose a sample fraction in order to achieve certain asymptotic optimality such as the minimal asymptotic mean squared error.

2.1 Optimal Sample Fractions in Estimation

1. Hill estimator In order to demonstrate how to bootstrap the optimal sample fraction in extreme-value statistics, we first consider estimation of the tail index in a special case when $\gamma > 0$. It is known that in this case, Eq. 1.1 is equivalent to

$$1 - F(x) = x^{-1/\gamma} \ell(x), \quad \text{for } x > 0, \tag{2.1}$$

where $\ell(x)$ is a slowly varying function with $\lim_{t\to\infty} \ell(tx)/\ell(t) = 1$ for all $x > 0$.
For $i \geq 1$ define

$$M^{(i)}(n, k) = \frac{1}{k} \sum_{i=1}^{k} (\log X_{n,n-i+1} - \log X_{n,n-k})^i.$$

Then

$$\hat{\gamma}^{(1)}(n, k) := M^{(1)}(n, k) = \frac{1}{k} \sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k}$$

is the well-known Hill estimator for γ proposed by Hill (1975), where $k = k(n)$
satisfies

$$k \to \infty \text{ and } k/n \to 0 \text{ as } n \to \infty. \tag{2.2}$$

Mason (1982) showed that $\hat{\gamma}^{(1)}(n, k)$ is a consistent estimator of γ.

In the study of the asymptotic distribution of $\hat{\gamma}^{(1)}(n, k)$, a popular second
order condition stronger than Eq. 2.1 has been used in many papers. Assume
that there exists a function $A(t)$ not changing sign at infinity such that

$$\lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \tag{2.3}$$

for all $x > 0$, where $\rho < 0$. In this case, $|A|$ is a regularly varying function
with index ρ (write $|A| \in R_\rho$); for details see de Haan and Stadtmüller (1996).
Under Eq. 2.3, one can write

$$\hat{\gamma}^{(1)}(n, k) - \gamma = \frac{\gamma}{\sqrt{k}} P_n^{(1)} + \frac{1}{1-\rho} \left(\frac{n}{k}\right) + o_p\left(A\left(\frac{n}{k}\right)\right), \tag{2.4}$$

where $P_n^{(1)}$ is a random variable which converges in distribution to a standard
normal distribution; see, e.g., de Haan and Peng (1998). Therefore, the as-
ymptotic mean squared error of $\hat{\gamma}^{(1)}(n, k)$, denoted by AMSE($\hat{\gamma}^{(1)}(n, k)$), is
given by

$$\text{AMSE}(\hat{\gamma}^{(1)}(n, k)) = \frac{\gamma^2}{k} + \frac{1}{(1-\rho)^2} A^2\left(\frac{n}{k}\right). \tag{2.5}$$

The optimal sample fraction for Hill's estimator is denoted by

$$k_1(n) = \arg\min_k \text{AMSE}(\hat{\gamma}^{(1)}(n, k)).$$

Our objective is to provide a consistent estimator for $k_1(n)$.

In Eq. 2.4, the Hill estimator is expanded after centered at γ. Since γ is unknown, a more suitable statistic for the bootstrap is $M(n, k) := M^{(2)}(n, k) - 2(M^{(1)}(n, k))^2$, which has a similar expansion:

$$M(n, k) = \frac{2\gamma^2}{\sqrt{k}} P_n^{(2)} + \frac{2\gamma\rho}{(1 - \rho)^2} A\left(\frac{n}{k}\right)(1 + o_p(1)), \tag{2.6}$$

where $P_n^{(2)}$ is a random variable converging in distribution to a standard normal distribution. Therefore, the asymptotic mean squared error for $M(n, k)$ is

$$\text{AMSE}(M(n, k)) = \frac{4\gamma^4}{k} + \frac{4\gamma^2\rho^2}{(1 - \rho)^4} A^2\left(\frac{n}{k}\right). \tag{2.7}$$

Set

$$\bar{k}_1(n) = \arg\min_k \text{AMSE}(M(n, k)).$$

Then $\bar{k}_1(n)$ and $k_1(n)$ are of the same order.

In order to determine the optimal sample fractions, we summarize a theorem from the proofs in de Haan and Peng (1998) or Danielsson et al. (2001).

Theorem 2.1 *For* $|A| \in R_\rho$ *($\rho < 0$), constants $\sigma > 0$ and $d \neq 0$, define*

$$k_0(n) = \arg\min_k \left(\frac{\sigma^2}{k} + d^2 A^2\left(\frac{n}{k}\right)\right).$$

Then

$$k_0(n) = \frac{n}{s^-(\sigma^2/(nd^2))}(1 + o(1)) \in R_{-2\rho/(1-2\rho)}, \quad \text{as } n \to \infty,$$

where s^- is the inverse function of s, with s given by

$$A^2(t) = \int_t^\infty s(u)du(1 + o(1)) \text{ as } t \to \infty.$$

From Eqs. 2.5 and 2.7 and Theorem 2.1 one can conclude the following theorem.

Theorem 2.2 (Danielsson et al. 2001) *Suppose Eq. 2.3 holds. Then*

$$k_1(n) = \frac{n}{s^-(\gamma^2(1 - \rho)^2/n)}(1 + o(1)),$$

$$\bar{k}_1(n) = \frac{n}{s^-(\gamma^2(1 - \rho)^4/(n\rho^2))}(1 + o(1)),$$

and therefore

$$\frac{\bar{k}_1(n)}{k_1(n)} \sim \left(1 - \frac{1}{\rho}\right)^{1/(1-2\rho)}. \tag{2.8}$$

Equation 2.8 indicates that estimating $k_1(n)$ is equivalent to estimating $\bar{k}_1(n)$ if a consistent estimator for the second order parameter ρ is available.

Set $\mathcal{X}_n = (X_1, \cdots, X_n)$. For a bootstrap sample $(X_1^*, \cdots, X_m^*)$, define the following bootstrap statistics

$$M_*^{(i)}(m, r) = \frac{1}{r} \sum_{i=1}^{r} \left(\log X_{m,m-i+1}^* - \log X_{m,m-r}^* \right)^i, \quad i = 1, 2 \qquad (2.9)$$

and $M_*(m, r) = M_*^{(2)}(m, r) - 2(M_*^{(1)}(m, r))^2$, where $r \to \infty$ and $r/m \to 0$, and $m = O(n^{1-\varepsilon})$ for some $0 < \varepsilon < 1$.

Conditional on $\mathcal{X}_n$, an expansion for $M_*(m, r)$, similar to Eq. 2.6, can be obtained. Let $r^*(m)$ be the integer such that the conditional second moment $E((M_*(m, r))^2 | \mathcal{X}_n)$ is minimal with respect to r. Then $r^*(m)$ can obtained by minimizing the estimated second moment $E((M_*(m, r))^2 | \mathcal{X}_n)$ via taking a large number of bootstrap samples. In fact, $r^*(m)$ has the following asymptotic property.

Theorem 2.3 (Danielsson et al. 2001) *Assume Eq. 2.3 holds. Then*

$$\frac{r^*(m) s^-(\gamma^2 (1 - \rho)^4 / (m \rho^2))}{m} \xrightarrow{p} 1 \ \ as \ n \to \infty.$$

The above two theorems play an important role in constructing a consistent bootstrap estimator for $\bar{k}_1(n)$. Now assume Eq. 2.3 holds with $A(t) = ct^\rho$. Then one can choose $s(t) = -2\rho c^2 t^{2\rho-1}$, and thus $s^-(u) = (2\rho c^2)^{1/(1-2\rho)} u^{-1/(1-2\rho)}$. It follows from Theorems 2.2 and 2.3 that

$$\left(\frac{m}{n} \right)^{2\rho/(1-2\rho)} \frac{r^*(m)}{\bar{k}_1(n)} \xrightarrow{p} 1 \ \ as \ n \to \infty. \qquad (2.10)$$

Although the second order parameter ρ can be estimated consistently by some estimator, say $\hat{\rho}_n$, it is not a good idea to use Eq. 2.10 directly to estimate $\bar{k}_1(n)$. This is due to the fact that it requires a faster convergence rate than $(\log n)^{-1}$ for $\hat{\rho}_n$ so that

$$\left(\frac{m}{n} \right)^{2\hat{\rho}_n/(1-2\hat{\rho}_n)} \Big/ \left(\frac{m}{n} \right)^{2\rho/(1-2\rho)} \xrightarrow{p} 1.$$

Fortunately, this obstacle can be overcome by obtaining two sub-sample bootstrap estimators, and the procedure is described as follows.

Step 1.   Set $m = m_1 = [n^{1-\varepsilon}]$ for some $\varepsilon \in (0, 1/2)$, where $[x]$ denotes the integer part of x. Draw a large number of bootstrap samples of size m from the empirical distribution F_n so that $E(M_*(m_1, r))^2 | \mathcal{X}_n)$ is well estimated and determine $r^*(m_1)$ such that the estimated $E(M_*(m_1, r))^2 | \mathcal{X}_n)$ is minimal at $r = r^*(m_1)$.

Step 2. Set $m = m_2 = [m_1^2/n]$, and repeat Step 1 to get $r^*(m_2)$ such that $E(M_*(m_2, r))^2|\mathcal{X}_n)$ is minimal at $r = r^*(m_2)$.

Step 3. Estimate ρ by $\hat{\rho}_n = \log r^*(m_1)/(-2\log m_1 + 2\log r^*(m_1))$, which is consistent.

Step 4. Define an estimator for $k_1(n)$ by

$$\hat{k}_1(n) := \frac{(r^*(m_1))^2}{r^*(m_2)}\left(1 - \frac{1}{\hat{\rho}_n}\right)^{1/(2\hat{\rho}_n - 1)}.$$

The consistency of the estimator follows from Eqs. 2.8 and 2.10.

Remark 1 Note that in the above procedure $m_1 = [n^{1-\varepsilon}]$ is an appropriate choice for any $\varepsilon \in (0, 1/2)$. A fully automatic procedure for the determination of m_1 is given in Danielsson et al. (2001); namely, choose m_1 such that the ratio

$$R(m_1) := \frac{\left(E(M_*(m_1, r^*(m_1)))^2|\mathcal{X}_n)\right)^2}{E(M_*(m_2, r^*(m_2)))^2|\mathcal{X}_n)}$$

is minimal.

Remark 2 We have seen that the validity of the sub-sample bootstrap procedure requires that $m_i \to \infty$ and $m_i/n \to 0$ for $i = 1, 2$ as $n \to \infty$. For a very large sample size n, this shouldn't be a problem. For a small or moderate sample, it might cause a problem since m_2 may be too small. For example, assume $n = 500$. If $m_1 = 100$, then $m_2 = [100^2/500] = 20$, and the number of observations used to construct the bootstrap statistics from a bootstrap sample is very small. Noting that the second bootstrap sample size $m_2 = [m_1^2/n]$ is selected in a way that one can cancel the factor in Eq. 2.10, a possible solution is to use the following formula $m_2 = [m_1(m_1/n)^a]$ for some fixed $a \in (0, 1]$. If a is small, this results in a larger m_2. Consequently, we have the following consistent estimator for $k_1(n)$:

$$\hat{k}_1^{(a)}(n) := \frac{(r^*(m_1))^{(a+1)/a}}{(r^*(m_2))^{1/a}}\left(1 - \frac{1}{\hat{\rho}_n}\right)^{1/(2\hat{\rho}_n - 1)}.$$

In particular, when $a = 1$, $\hat{k}_1^{(1)}(n) = \hat{k}_1(n)$.

2. Moment estimator Let $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in R$. The moment estimator for γ, proposed by Dekkers et al. (1989), is defined as

$$\hat{\gamma}^{(2)}(n, k) = M^{(1)}(n, k) + 1 - \frac{1}{2}\left(1 - \frac{(M^{(1)}(n, k))^2}{M^{(2)}(n, k)}\right)^{-1}.$$

Set $\gamma_+ = \max(\gamma, 0)$ and $\gamma_- = \min(\gamma, 0)$. The assumption $F \in \mathcal{D}(G_\gamma)$ is equivalent to the condition that there is a positive function $a(t)$ such that

$$\lim_{t\to\infty} \frac{U(at) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \text{ for } x > 0,$$

which implies

$$\lim_{t \to \infty} \frac{\log U(at) - \log U(t)}{a(t)/U(t)} = \frac{x^{\gamma_-} - 1}{\gamma_-} \quad \text{for } x > 0.$$

We need the following second order condition: there exists a function $A(t)$ not changing sign at infinity such that

$$\lim_{t \to \infty} \frac{\frac{\log U(at) - \log U(t)}{a(t)/U(t)} - \frac{x^{\gamma_-} - 1}{\gamma_-}}{A(t)} = H(x), \tag{2.11}$$

where

$$H(x) = \frac{1}{\rho} \left(\frac{x^{\rho + \gamma_-} - 1}{\rho + \gamma_-} - \frac{x^{\gamma_-} - 1}{\gamma_-} \right)$$

$$= \begin{cases} (\log x)^2/2 & \text{if } \rho = 0, \gamma \geq 0, \\ \dfrac{1}{\gamma} \left(x^\gamma \log x - \dfrac{x^\gamma - 1}{\gamma} \right) & \text{if } \rho = 0, \gamma < 0, \\ \dfrac{1}{\rho} \left(\dfrac{x^{\rho + \gamma_-} - 1}{\rho + \gamma_-} - \dfrac{x^{\gamma_-} - 1}{\gamma_-} \right) & \text{if } \rho < 0. \end{cases}$$

Similarly to the Hill estimator, $\hat{\gamma}^{(2)}(n, k) - \gamma$ can be decomposed into two parts, a non-degenerate term with an asymptotic normal distribution and a degenerate bias term, and the asymptotic mean squared error of $\hat{\gamma}^{(2)}(n, k)$ is given by

$$\text{AMSE}(\hat{\gamma}^{(2)}(n, k)) = \frac{V^2(\gamma)}{k} + b^2(\gamma, \rho) A_0^2 \left(\frac{n}{k} \right),$$

where

$$V^2(\gamma) = \begin{cases} \gamma^2 + 1 & \text{if } \gamma > 0, \\ \dfrac{(1 - \gamma)^2 (1 - 2\gamma)(6\gamma^2 - \gamma + 1)}{(1 - 3\gamma)(1 - 4\gamma)} & \text{if } \gamma < 0, \end{cases}$$

$$b(\gamma, \rho) = \begin{cases} \dfrac{\gamma}{\rho(1 - \rho)} + \dfrac{1}{(1 - \rho)^2} & \text{if } \gamma > 0, \\ \dfrac{1}{1 - \gamma} & \text{if } \rho < \gamma < 0, \\ \dfrac{(1 - \gamma)(1 - 2\gamma)}{(1 - \rho - \gamma)(1 - \rho - 2\gamma)} & \text{if } \gamma < \rho \end{cases}$$

and

$$A_0(t) = \begin{cases} A(t) & \text{if } \gamma > 0, \\ \dfrac{a(t)}{U(t)} & \text{if } \rho < \gamma < 0, \\ A(t) & \text{if } \gamma < \rho. \end{cases}$$

Theorem 2.4 (Draisma et al. 1999) *Suppose that $F \in \mathcal{D}(G_\gamma)$ and that Eq. 2.11 holds for $\rho < 0$, $\gamma \neq \rho$ and $\gamma \neq 0$. Set $k_2(n) = \arg\min_k \mathrm{AMSE}(\hat{\gamma}^{(2)}(n, k))$. Then*

$$k_2(n) \sim \left(\frac{V^2(\gamma)}{b^2(\gamma, \rho)} \right)^{1/(1-2\rho^*)} \frac{n}{s^-\left(\frac{1}{n}\right)}$$

as $n \to \infty$, where s^- is the inverse function of the decreasing function s satisfying

$$A_0^2(t) = \int_t^\infty s(u)du(1 + o(1)).$$

Define an alternative estimator for γ by

$$\hat{\gamma}^{(3)}(n, k) = \sqrt{M^{(2)}(n, k)/2} + 1 - \frac{2}{3}\left(1 - \frac{M^{(1)}(n, k)M^{(2)}(n, k)}{M^{(3)}(n, k)}\right)^{-1}.$$

Then the auxiliary statistic $\hat{\gamma}^{(2)}(n, k) - \hat{\gamma}^{(3)}(n, k)$ has a similar asymptotic behavior to that of $\hat{\gamma}^{(2)}(n, k) - \gamma$.

Theorem 2.5 (Draisma et al. 1999) *Assume that the conditions of Theorem 2.4 hold. Set*

$$\bar{k}_2(n) = \arg\min_k \mathrm{AMSE}(\hat{\gamma}^{(2)}(n, k) - \hat{\gamma}^{(3)}(n, k)).$$

Then

$$\bar{k}_2(n) \sim \left(\frac{\bar{V}^2(\gamma)}{\bar{b}^2(\gamma, \rho)} \right)^{1/(1-2\rho^*)} \frac{n}{s^-\left(\frac{1}{n}\right)} \quad as \ n \to \infty,$$

where

$$\bar{V}^2(\gamma) = \begin{cases} \dfrac{1}{4}(\gamma^2+1) & \text{if } \gamma > 0, \\[2mm] \dfrac{1}{4}\dfrac{(1-\gamma)^2(1-8\gamma+48\gamma^2-154\gamma^3+263\gamma^4-222\gamma^5+72\gamma^6)}{(1-2\gamma)(1-3\gamma)(1-4\gamma)(1-5\gamma)(1-6\gamma)} & \text{if } \gamma < 0, \end{cases}$$

$$\bar{b}(\gamma, \rho) = \begin{cases} -\dfrac{\gamma(1-\rho)+\rho}{2(1-\rho)^3} & \text{if } \gamma > 0, \\[2mm] \dfrac{(1-2\rho)-\sqrt{(1-\rho)(1-2\rho)}}{(1-\rho)(1-2\rho)} & \text{if } \rho < \gamma < 0, \\[2mm] \dfrac{1}{2}\dfrac{-\rho(1-\gamma)^2}{(1-\rho-\gamma)(1-\rho-2\gamma)(1-\rho-3\gamma)} & \text{if } \gamma < \rho. \end{cases}$$

For a bootstrap sample of size m, $(X_1^*, \cdots, X_m^*)$, define $M_*^{(i)}(m, r)$ as in Eq. 2.9 and set

$$\hat{\gamma}_*^{(2)}(m, r) = M_*^{(1)}(m, r) + 1 - \frac{1}{2}\left(1 - \frac{(M_*^{(1)}(m, r))^2}{M_*^{(2)}(m, r)}\right)^{-1},$$

and

$$\hat{\gamma}_*^{(3)}(m, r) = \sqrt{M_*^{(2)}(m, r)/2} + 1 - \frac{2}{3}\left(1 - \frac{M_*^{(1)}(m, r) M_*^{(2)}(m, r)}{M_*^{(3)}(m, r)}\right)^{-1},$$

where $r \to \infty$ and $r/m \to 0$, and $m = O(n^{1-\varepsilon})$ for some $0 < \varepsilon < 1$. Further, set for some $\delta > 0$

$$\hat{\gamma}_*^{(4)}(m, r) = \left(\hat{\gamma}_*^{(2)}(m, r) - \hat{\gamma}_*^{(3)}(m, r)\right) I\left(|\hat{\gamma}_*^{(2)}(m, r) - \hat{\gamma}_*^{(3)}(m, r)| < r^{\delta-1/2}\right).$$

Theorem 2.6 (Draisma et al. 1999) *Suppose the conditions of Theorem 2.4 hold and $m = O(n^{1-\varepsilon})$ for some $0 < \varepsilon < 1/2$. Define $r_3^*(m) = \arg\min_r E((\hat{\gamma}_*^{(4)}(m, r))^2 | \mathcal{X}_n)$. Then*

$$r_3^*(m) \bigg/ \left\{\left(\frac{\bar{V}^2(\gamma)}{\bar{b}^2(\gamma, \rho)}\right)^{1/(1-2\rho^*)} \frac{m}{s^-(\frac{1}{m})}\right\} \xrightarrow{p} 1 \ as \ n \to \infty.$$

Suppose the conditions of Theorem 2.6 hold and $A_0(t) = ct^{\rho^*}$. Then from Theorems 2.5 and 2.6 we have

$$\frac{\bar{k}_2(n)}{r_3^*(m)}\left(\frac{m}{n}\right)^{\frac{-2\rho^*}{1-2\rho^*}} \xrightarrow{p} 1.$$

Using the same procedure for Hill's estimator, we obtain $r_3^*(m_1)$ and $r_3^*(m_2)$, where $m_2 = [m_1^2/n]$. Define

$$\hat{k}_2(n) = \frac{(k_3^*(m_1))^2}{k_3^*(m_2)}\left(\frac{V^2(\hat{\gamma}_n)\bar{b}^2(\hat{\gamma}_n, \hat{\rho}_n^*)}{\bar{V}^2(\hat{\gamma}_n)b^2(\hat{\gamma}_n, \hat{\rho}_n^*)}\right)^{\frac{1}{1-\hat{\rho}_n^*}},$$

where $\hat{\gamma}_n$ and $\hat{\rho}_n^*$ are any consistent estimators of γ and ρ^*, respectively. Then

$$\frac{\hat{k}_2(n)}{k_2(n)} \xrightarrow{p} 1.$$

3. More applications We have introduced two novel applications of the bootstrap method. In fact, under certain technical conditions, the bootstrap approach has proved to be valid in estimating the optimal sample fractions for Pickands-type estimators of the tail index (Draisma et al. 1999), estimators of high quantiles and endpoints (Ferreira et al. 2003), and estimators of small exceedances probability (Ferreira 2002). The bootstrap procedures for these estimators are the same. Below is a summary of the fundamental steps used to estimate the optimal sample fraction, say $k_0(n)$, that minimizes the asymptotic mean squared error of an estimator of the aforementioned characteristics.

- Construct an auxiliary statistic in a way that $\bar{k}_0(n)/k_0(n) \sim C$, where $\bar{k}_0(n)$ is the sample fraction that minimizes the asymptotic second moment of

the auxiliary statistic, and C is a constant depending on some unknown parameters and can be consistently estimated by a statistic, say $\hat{C}$.

- Use Step 1 and Step 2 introduced for Hill's estimator to estimate two optimal sample fractions that minimize the second moments of the two bootstrap versions of the auxiliary statistic based on two bootstrap samples of sizes m_1 and m_2, and denote them as $r^*(m_1)$ and $r^*(m_2)$ accordingly.

- Define $\hat{k}_0(n) = \dfrac{(r^*(m_1))^2}{\hat{C}r^*(m_2)}$. Then $\hat{k}_0(n)$ is a consistent estimator of $k_0(n)$, that is, $\dfrac{\hat{k}_0(n)}{k_0(n)} \overset{p}{\to} 1$.

For a class of kernel estimators of the tail index, the selection of the bandwidth is parallel to the selection of the sample fraction in the above estimators. A similar bootstrap procedure for optimal bandwidth choice for kernel estimators can be found in Groeneboom et al. (2003).

Pictet et al. (1998) derived an algorithm to reduce the bias of the Hill estimator. The algorithm is based on a subsample bootstrap combined with the jackknife method.

It is worth mentioning that some methods other than the bootstrap can be found in the literature. For example, Drees and Kaufmann (1998) and Matthys and Beirlant (2000) proposed some automatic methods for moment-type estimators of the tail index, and Fraga Alves (2001) gave an algorithm for estimation of the optimal sample fraction in the location invariant Hill-type estimators.

2.2 Bootstrap Confidence Intervals

The bootstrap method has been commonly used to estimate the sampling distribution of a statistic which allows us to build confidence intervals of a population characteristic. The bootstrap approach avoids estimating the unknown parameters in the limiting distribution of the statistic, and it is extremely useful when the limiting distribution of the statistic is unknown or too complicated to calculate. In many situations such an approach is satisfactory in the sense that the bootstrap confidence intervals have a good coverage accuracy.

Guillou (2000) assessed the accuracy of the sub-sample bootstrap method for the tail index of a heavy-tailed distribution and found out that the accuracy of the sub-sample bootstrap approach is often worse than that of the asymptotic approximation of the statistic. Bacro and Brito (1998) proposed a different bootstrap procedure for the tail index of a heavy-tailed distribution by resampling directly from $k = k_n$ log-spacings, $i(\log X_{n,n-i+1} - \log X_{n,n-i})$, $i = 1, \ldots, k$. They didn't consider the accuracy of the coverage for the approach. Nevertheless, this procedure is valid if the Hill estimator is asymptotically unbiased.

When applying the sub-sample bootstrap procedure, one has to decide how to choose the size of the sub-samples. Caers and Van Dyck (1998) and Guillou (2000) discussed this issue. Caers and Van Dyck (1998) even proposed a double

bootstrap method that estimates the optimal sample fraction based on a sub-sample bootstrap in the first stage and then constructs the confidence intervals based on a sub-sample bootstrap from the original data in the second stage.

On the other hand, a full-sample bootstrap still works when the sample fraction in the estimation is in the range that the estimator is asymptotically unbiased. El-Nouty and Guillou (2000) examined an estimator for the tail index of a heavy-tailed distribution and proved the validity of the full-sample bootstrap. Peng and Qi (2007) applied a full-sample bootstrap method to obtain the confidence band for the tail dependence function in bivariate extremes. Both papers obtained the accuracy of the bootstrap method by using an approach proposed by Chen and Lo (1997). In Chen and Lo's approach, the key is to construct a statistic such that it is identically distributed as the statistic of interest and it is also very close to the bootstrapped statistic.

3 Empirical Likelihood Methods

Empirical likelihood was introduced by Owen (1988, 1990) for a mean vector for iid observations. It is a nonparametric method that allows one to employ likelihood methods to construct confidence intervals without assuming a parametric family for the data. It produces confidence regions whose shape and orientation are determined entirely by the data. Just like the parametric likelihood method, the method involves only a maximization procedure, and the estimation of any nuisance parameter is not necessary. Therefore, the empirical likelihood method possesses some advantages of the likelihood method.

There is a large and growing literature extending empirical likelihood methods to many statistical problems. For recent references and development we refer to the book by Owen (2001).

The empirical likelihood method was first adopted by Lu and Peng (2002) to construct confidence intervals for the tail index of a heavy-tailed distribution. Peng and Lu applied both the empirical likelihood method and the (pseudo) parametric likelihood method to obtain confidence intervals for the tail index of a heavy-tailed distribution. Later, Peng and Qi (2006a) proposed a new calibration method to overcome the undercoverage problem in constructing confidence intervals. Peng (2004) applied the empirical likelihood method to construct the confidence intervals for a mean from a heavy-tailed distribution. Peng and Qi (2006b) employed a data tilting method to construct the confidence intervals (or regions) for the high quantiles of a heavy-tailed distribution. The data tilting method is a general method and is more flexible compared with the empirical likelihood method. It has been demonstrated in Peng and Qi (2006a, b) that the empirical likelihood method and the data tilting method outperform the normal approximation method in that both the empirical likelihood method and the data tilting method result in confidence intervals with shorter average length and more accurate coverage probability.

Here we assume that F is a heavy-tailed distribution satisfying the second-order condition (2.3).

3.1 Index of a Heavy-Tailed Distribution

Define $Y_i = i(\log X_{n,n-i+1} - \log X_{n,n-i})$ for $i = 1, \ldots, k$. Then $\{Y_i, 1 \le i \le k\}$ are asymptotically independent with a common exponential limiting distribution with mean γ for fixed k (see e.g., Weissman 1978). Lu and Peng (2002) applied Owen's empirical likelihood method to construct confidence interval for the index γ. The procedure is as follows.

Let $p = (p_1, \cdots, p_k)$ be a probability vector satisfying $\sum_{i=1}^{k} p_i = 1$ and $p_i \ge 0$ for $i = 1, \ldots, k$. Then the empirical likelihood, evaluated at true value γ_0 for the tail index γ, is defined by

$$L(\gamma_0) = \sup \left\{ \prod_{i=1}^{k} p_i : \sum_{i=1}^{k} p_i Y_i = \gamma_0 \right\}.$$

Then, by the method of Lagrange multipliers, one can show that

$$p_i = \frac{1}{k} \{1 + \lambda(Y_i - \gamma_0)\}^{-1}, \quad i = 1, \ldots, k,$$

where λ is the solution to the equation

$$\frac{1}{k} \sum_{i=1}^{k} \frac{Y_i - \gamma_0}{1 + \lambda(Y_i - \gamma_0)} = 0. \tag{3.1}$$

Note that $\prod_{i=1}^{k} p_i$, subject to $\sum_{i=1}^{k} p_i = 1$, achieves its maximum k^{-k} at $p_i = k^{-1}$. Thus, the empirical likelihood ratio at γ_0 is given by

$$l(\gamma_0) = \prod_{i=1}^{k} (kp_i) = \prod_{i=1}^{k} \{1 + \lambda(Y_i - \gamma_0)\}^{-1},$$

and the corresponding empirical log-likelihood ratio statistic is defined as

$$\mathcal{L}(\gamma_0) = -2 \log l(\gamma_0) = 2 \sum_{i=1}^{k} \log\{1 + \lambda(Y_i - \gamma_0)\},$$

where λ is the solution of Eq. 3.1.

Theorem 3.1 (Lu and Peng 2002) *Under condition Eq. 2.3, if*

$$k = k_n \to \infty, \quad \frac{k}{n} \to 0 \text{ and } \sqrt{k} A(n/k) \to 0 \text{ as } n \to \infty,$$

then

$$\mathcal{L}(\gamma_0) \xrightarrow{d} \chi_1^2,$$

where χ_1^2 denotes a chi-squared random variable with one degree of freedom, and γ_0 is the true value of the tail index γ.

According to Theorem 3.1, a $100(1 - \alpha)\%$ confidence interval for γ based on the empirical likelihood ratio statistic is determined by

$$I_E(1 - \alpha) = \{\gamma > 0 : \mathcal{L}(\gamma) \leq c(\alpha)\},$$

where $c(\alpha)$ is the α level critical value of a chi-squared distribution with one degree of freedom.

For a small sample size, the asymptotic χ^2 calibrated empirical likelihood-based confidence regions for iid observations may have a lower coverage probability than the nominal level, as indicated by numerical evidence in the literature; see, e.g., Owen (1988), Hall and La Scala (1990), and Qin and Lawless (1994). The reason for the undercoverage was explained in Tsao (2004), that is, the distribution of the empirical likelihood ratio has an atom at infinity, and the atom can be substantial if the sample size is not large. For the empirical likelihood-based confidence interval for the tail index the same problem occurs since only a few upper order statistics are employed in the inference. As a remedy for the empirical likelihood method, Peng and Qi (2006a) proposed the following new calibration method to improve the coverage accuracy.

Since Y_i/γ in $\mathcal{L}(\gamma)$ has an approximate exponential distribution with mean one, we replace $Y_1/\gamma, \ldots, Y_k/\gamma$ by i.i.d. random variables $E_1, \ldots, E_k$ with an exponential distribution with mean one. Define

$$\text{ELR}(k) = 2 \sum_{i=1}^{k} \log(1 + \lambda'(E_i - 1)),$$

where λ' is the solution to

$$\sum_{i=1}^{k} \frac{E_i - 1}{1 + \lambda'(E_i - 1)} = 0.$$

Then approximate the distribution of $\mathcal{L}(\gamma)$ by the distribution of $\text{ELR}(k)$ instead of χ_1^2. Therefore, the following confidence interval with nominal level $100(1 - \alpha)\%$ can be defined:

$$I_E^*(1 - \alpha) = \{\gamma : \mathcal{L}(\gamma) \leq c(k, \alpha)\},$$

where $c(k, \alpha)$ is the upper α-level critical value of the distribution of $\text{ELR}(k)$.

Note that the exact distribution of $\text{ELR}(k)$ is not available. The critical value $c(k, \alpha)$ can be estimated via Monte Carlo simulation. Peng and Qi (2006a) obtained the critical values $c(k, \alpha)$ for $\alpha = 10\%, 5\%$ and 1% for all k between 10 and 200 based on 1 000 000 random samples. These critical values for $10 \leq k \leq 29$ were listed in Table 1 in Peng and Qi (2006a), and for $30 \leq k \leq 200$, three linear regression equations are fitted by

$$c(k, 0.10) = 2.7055 - \frac{0.51269}{\sqrt{k}} + \frac{18.14242}{k},$$

$$c(k, 0.05) = 3.8415 - \frac{1.12486}{\sqrt{k}} + \frac{32.90613}{k}$$

and

$$c(k, 0.01) = 6.6349 - \frac{4.56941}{\sqrt{k}} + \frac{98.98899}{k}.$$

The three intercepts in the equations are the 10%, 5% and 1% critical values of the χ_1^2 distribution.

The simulation study in Peng and Qi (2006a) indicates that this new calibration method greatly improves the accuracy of the coverage probability for empirical likelihood-based confidence intervals for small sample proportion k. Numerically it also demonstrates that the empirical likelihood method outperforms both the normal approximation (cf., de Haan and Peng 1998) and the gamma approximation (cf., Cheng and de Haan 2001) of Hill's estimators in terms of coverage probability and length of confidence intervals.

3.2 High Quantile

Recall that x_p is called a high quantile if $p = p_n \in (0, 1)$ is very small in the sense that $np_n \to d \in [0, \infty)$. Peng and Qi (2006b) employed a data tilting method to construct a confidence interval for x_p.

To introduce the methodology, we assume temporarily that $1 - F(x) = cx^{-1/\gamma}$ for $x > T$, where $T > 0$ is a large number. Set $\delta_i = I(X_i > T)$. Then the likelihood function for the censored data $\{(\delta_i, \max(X_i, T))\}_{i=1}^n$ is $\Pi_{i=1}^n (c\gamma^{-1} X_i^{-1/\gamma-1})^{\delta_i} (1 - cT^{-1/\gamma})^{1-\delta_i}$. Below we take $T = X_{n,n-k}$.

First, for any fixed weights $q = (q_1, \cdots, q_n)$ such that $q_i \geq 0$ and $\sum_{i=1}^n q_i = 1$, define a weighted log-likelihood function by

$$L(\gamma, c) = \sum_{i=1}^n q_i \log \left((c\gamma^{-1} X_i^{-1/\gamma-1})^{\delta_i} (1 - cX_{n,n-k}^{-1/\gamma})^{1-\delta_i} \right).$$

We solve $(\hat{\gamma}(q), \hat{c}(q)) = \arg\max_{(\gamma,c)} L(\gamma, c)$ and obtain

$$\begin{cases} \hat{\gamma}(q) = \dfrac{\sum_{i=1}^n q_i \delta_i (\log X_i - \log X_{n,n-k})}{\sum_{i=1}^n q_i \delta_i} \\ \hat{c}(q) = X_{n,n-k}^{1/\hat{\gamma}(q)} \sum_{i=1}^n q_i \delta_i. \end{cases}$$

Define $D(q) = \sum_{i=1}^n q_i \log(nq_i)$, which is a measure of distance between q and the uniform distribution, i.e. $q_i = 1/n$. Next, we shall choose q to minimize this distance. More specifically, solve $(2n)^{-1} L(x_p) = \min_q D(q)$ subject to the constraints

$$q_i \geq 0, \ \sum_{i=1}^n q_i = 1, \log \frac{x_p}{X_{n,n-k}} = \hat{\gamma}(q) \log \frac{\sum_{i=1}^n q_i \delta_i}{p_n}. \tag{3.2}$$

Put

$$A_1(\lambda_1) = 1 - \frac{n-k}{n} e^{-1-\lambda_1} \quad \text{and} \quad A_2(\lambda_1) = A_1(\lambda_1) \frac{\log(x_p/X_{n,n-k})}{\log(A_1(\lambda_1)/p_n)}.$$

Then, by the standard method of Lagrange multipliers, we have

$$q_i = q_i(\lambda_1, \lambda_2) = \begin{cases} \dfrac{1}{n} e^{-1-\lambda_1}, \text{ if } \delta_i = 0, \\[3mm] \dfrac{1}{n} \exp\left\{ -1 - \lambda_1 + \lambda_2 \left(\dfrac{\log(x_p/X_{n,n-k})}{A_2(\lambda_1)} - \dfrac{1}{A_1(\lambda_1)} \right. \right. \\[3mm] \left. \left. - \dfrac{A_1(\lambda_1) \log(X_i/X_{n,n-k}) \log(x_p/X_{n,n-k}))}{A_2^2(\lambda_1)} \right) \right\}, \text{ if } \delta_i = 1, \end{cases}$$

where λ_1 and λ_2 satisfy Eq. 3.2.

Theorem 3.2 (Peng and Qi 2006b) *Assume Eqs. 2.3 and 2.2 hold. If $\sqrt{k}A(n/k) \to 0$, $np_n = O(k)$ and $\log\left(\frac{k}{np_n}\right)/\sqrt{k} \to 0$ as $n \to \infty$, then, with probability tending to one, there exists a solution to Eq. 3.2, and $L(x_{p,0}) \xrightarrow{d} \chi_1^2$, where $x_{p,0}$ is the true value of x_p.*

According to the above theorem, a confidence region at level $1 - \alpha$ for x_p is given by

$$I_T(1-\alpha) = \{x_p : L(x_p) \le c(\alpha)\},$$

where $c(\alpha)$ is the α-level critical value of χ_1^2. This confidence interval has asymptotically correct coverage probability α, i.e., $P(x_p \in I_T(1-\alpha)) \to \alpha$ as $n \to \infty$.

References

Angus, J.E.: Asymptotic theory for bootstrapping the extremes. Commun. Stat. Theory Methods **22**, 15–30 (1993)

Bacro, J.N., Brito, M.: A tail bootstrap procedure for estimating the tail Pareto-index. J. Stat. Plan. Inference **71**, 245–260 (1998)

Caers, J., Van Dyck, J.: Nonparametric tail estimation using a double bootstrap method. Comput. Stat. Data Anal. **29**, 191–211 (1998)

Chen, K., Lo, S.H.: On a mapping approach to investigating the bootstrap accuracy. Probab. Theory Relat. Fields **107**, 197–217 (1997)

Cheng, S., de Haan, L. Penultimate approximation for Hill's Estimator. Scand. J. Statist. **28**, 569–575 (2001)

Danielsson, J., de Haan, L., Peng, L., de Vries, C.G.: Using a bootstrap method to choose the sample fraction in tail index estimation. J. Multivar. Anal. **76**, 226–248 (2001)

Danielsson, J., Vries, C.: Tail index and quantile estimation with very high frequency data. J. Empir. Finance **4**, 241–257 (1997)

de Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer, New York (2006)

de Haan, L., Peng, L.: Comparison of tail index estimators. Statist. Neerlandica **52**, 60–70 (1998)

de Haan, L., Stadtmüller, U.: Generalized regular variation of second order. J. Aust. Math. Soc. A **61**, 381–395 (1996)

Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L.: A moment estimator for the index of an extreme-value distribution. Ann. Stat. **17**, 1833–1855 (1989)

Draisma, G., de Haan, L., Peng, L., Pereira, T.T.: A bootstrap-based method to achieve optimality in estimating the extreme-value index. Extremes **2**, 367–404 (1999)

Drees, H., Kaufmann, E.: Selection of the optimal sample fraction in univariate extreme value estimation. Stoch. Process. Appl. **75**, 149–172 (1998)

Efron, N.: Bootstrap methods: another look at the jackknife. Ann. Stat. **7**, 1–26 (1979)

El-Nouty, C., Guillou, A.: On the bootstrap accuracy of the Pareto index. Stat. Decis. **18**, 275–290 (2000)

Ferreira, A.: Optimal asymptotic estimation of small exceedance probabilities. J. Stat. Plan. Inference **104**, 83–102 (2002)

Ferreira, A., de Haan, L., Peng, L.: On optimising the estimation of high quantiles of a probability distribution. Statistics **37**(5), 401–434 (2003)

Fraga Alves, M.I.: A location invariant Hill-type estimator. Extremes **4**(3), 199–217 (2001)

Galambos, J.: The asymptotic Theory of Extreme Order Statistics. Wiley, New York (1978)

Geluk, J., de Haan, L.: On bootstrap sample size in extreme value theory. Publ. Inst. Math. (Beograd) (N.S.) **71**(85), 21–25 (2002)

Groeneboom, P., Lopuhaä, H.P., de Wolf, P.P.: Kernel-type estimators for the extreme value index. Ann. Stat. **31**, 1956–1995 (2003)

Guillou, A.: Bootstrap confidence intervals for the Pareto index. Commun. Stat. Theory Methods **29**, 211–226 (2000)

Hall, P.: Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. J. Multivar. Anal. **32**, 177–203 (1990)

Hall, P., La Scala, L.: Methodology and algorithms of empirical likelihood. Internat. Statist. Rev. **58**, 109–127 (1990)

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

Lu, J., Peng, L.: Likelihood based confidence intervals for the tail index. Extremes **5**(4), 337–352 (2002)

Mason, D.M.: Laws of large numbers for sums of extreme values. Ann. Probab. **10**, 754–764 (1982)

Matthys, G., Beirlant, J.: Adaptive threshold selection in tail index estimation. In: Embrechts P. (ed.) Extremes and Integrated Risk Management, Risk Books, pp. 37–49. London (2000)

Owen, A.: Empirical likelihood ratio confidence intervals for single functional. Biometrika **75**, 237–249 (1988)

Owen, A.: Empirical likelihood regions. Ann. Stat. **18**, 90–120 (1990)

Owen, A.: Empirical Likelihood. Chapman and Hall (2001)

Peng, L.: Empirical likelihood confidence interval for a mean with a heavy tailed distribution. Ann. Stat. **32**, 1192–1214 (2004)

Peng, L., Qi, Y.: A new calibration method of constructing empirical likelihood-based confidence intervals for the tail index. Aust. N.Z. J. Stat. **48**, 59–66 (2006a)

Peng, L., Qi, Y.: Confidence regions for high quantiles of a heavy tailed distribution. Ann Stat. **34**(4), 1964–1986 (2006b)

Peng, L., Qi, Y.: Bootstrap approximation of tail dependence function. J. Multivar. Anal. (2007) (in press)

Pereira, T.T.: Second order behavior of domains of attraction and the bias of generalized Pickands' estimator. In: Lechner, J., Galambos, J., Simin, E. (eds.) Extreme Value Theory and Applications III. Proc. Gaithersburg Conference (NIST special publ) (1993)

Pickands III, J.: Statistical inference using extreme order statistics. Ann. Stat. **3**, 119–131 (1975)

Pictet, O.V., Dacorogna, M.M., Müller, U.A.: Hill, bootstrap and jackknife estimators for heavy tails. In: Adler, R., Feldman, R., Taqqu, M.S. (eds.) A Practical Guide to Heavy Tails, pp. 283–310. Birkhauser, Boston (1998)

Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. Ann. Stat. **22**, 300–325 (1994)

Shao, J., Tu, D.: The Jackknife and the Bootstrap. Springer (1995)

Tsao, M.: A new method of calibration for the empirical loglikelihood ratio. Statist. Probab. Lett. **68**, 305–314 (2004)

Weissman, I.: Estimation of parameters and large quantiles based on the k largest observations. J. Amer. Statist. Assoc. **73**, 812–815 (1978)