

# Bootstrapping Endpoint

Zhouping Li<sup>1</sup> and Liang Peng<sup>2</sup>

<sup>1</sup> School of Mathematics and Statistics, Lanzhou University,  
Lanzhou 730000, China lizhp@lzu.edu.cn

<sup>2</sup> School of Mathematics, Georgia Institute of Technology,  
Atlanta, 30332-0160, USA peng@math.gatech.edu

## Abstract

It is known that bootstrapping maximum for estimating the endpoint of a distribution function is inconsistent and subsample bootstrap method is needed. Under an extreme value condition, some other estimators for the endpoint have been studied in the literature, which are preferable to the maximum in regular cases. In this paper, we show that the full sample bootstrap method is consistent for the endpoint estimator proposed by Hall (1982).

**Keywords:** Bootstrap; endpoint; extreme-value index.

# 1 Introduction

Let  $X_1, \dots, X_n$  be independent and identically distributed random variables with distribution function  $F(x)$ , which lies in the domain of attraction of an extreme value distribution, i.e., there exist  $a_n > 0$  and  $b_n \in R$  such that

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) := \exp\{-(1 + \gamma x)^{-1/\gamma}\}, \quad (1)$$

where  $1 + \gamma x > 0$  and  $\gamma \in R$  is called the extreme value index. Denote the right endpoint of  $F$  by  $\theta$ , i.e.,  $\theta = \sup\{x : F(x) < 1\}$ . Under condition (1), it is known that  $\theta$  is finite in case of  $\gamma < 0$  and some studies on estimating the endpoint  $\theta$  in case of  $\gamma < 0$  exist in the literature; see Woodroffe (1974), Hall (1982), Loh (1984), Smith (1987), Dekker et al. (1989), Aarssen and de Haan (1994), Athreya and Fukuchi (1997), Hall and Wang (1999), Li and Peng (2009, 2010) and Peng and Qi (2009).

Assume that

$$1 - F(x) = c(\theta - x)^{-1/\gamma} \{1 + O((\theta - x)^\alpha)\} \quad \text{as } x \uparrow \theta, \quad (2)$$

where  $\alpha > 0$  and  $c > 0$ , Hall (1982) proposed the following maximum likelihood estimator

$$\begin{aligned} & (\hat{c}_{Hall}(k), \hat{\theta}_{Hall}(k), \hat{\gamma}_{Hall}(k)) \\ = & \arg \max \frac{n!}{(n-k-1)!} \left\{ -\frac{c}{\gamma} \right\}^{k+1} \left\{ \prod_{i=1}^{k+1} (\theta - X_{n,n-i+1})^{-1/\gamma-1} \right\} \times \\ & \left\{ 1 - c(\theta - X_{n,n-k})^{-1/\gamma} \right\}^{n-k-1}, \end{aligned}$$

which is equivalent to that  $\hat{\theta}_{Hall}(k)$  is the solution of the following equation

$$\frac{1}{k+1} \sum_{i=1}^k \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-i+1}} \left\{ \frac{1}{k+1} \sum_{i=1}^k \log \frac{\theta - X_{n,n-i+1}}{\theta - X_{n,n-k}} + 1 \right\} = 1 \quad (3)$$

and

$$\hat{\gamma}_{Hall}(k) = \frac{1}{k+1} \sum_{i=1}^k \log \frac{\hat{\theta}_{Hall}(k) - X_{n,n-i+1}}{\hat{\theta}_{Hall}(k) - X_{n,n-k}}. \quad (4)$$

Here  $X_{n,1} \leq \dots \leq X_{n,n}$  denote the order statistics of  $X_1, \dots, X_n$ . Moreover, Hall (1982) derived the asymptotic distribution of  $(\hat{\theta}_{Hall}(k), \hat{\gamma}_{Hall}(k))$  when  $\sqrt{k}(k/n)^{-\gamma\alpha} \rightarrow 0$ . Recently, assuming the second order condition (7) below and

$$k = k(n) \rightarrow \infty \quad \text{and} \quad \sqrt{k}A(n/k) \rightarrow \lambda \in R \quad \text{as } n \rightarrow \infty,$$

where  $A$  is defined in (7) below, Li and Peng (2010) extended the results in Hall (1982) by adding the bias terms of the joint asymptotic distribution of  $(\hat{\gamma}_{Hall}(k), \hat{\theta}_{Hall}(k))$ .

In this paper, we are interested in constructing a confidence interval for  $\theta$ . One way is to employ the normal approximation method based on the asymptotic distribution of  $\hat{\theta}_{Hall}(k)$ . Another commonly used approach in constructing confidence intervals is the bootstrap method. However, it is known that full sample bootstrap method is inconsistent for extremes. For example, Bickel and Freedman (1981) pointed out that the full sample bootstrapping distribution of the normalized maximum fails to be consistent when  $F$  is the uniform distribution, and Angus (1993) showed the same result when  $F$  lies in the domain

of attraction of an extreme value distribution. Instead, subsample bootstrap methods have been proposed to deal with extremes; see Swanepoel (1986), Deheuvels et al. (1993), Athreya and Fukuchi (1997). More references can be found in Qi (2008), which reviewed some applications of bootstrap methods in analyzing extreme values.

In this paper, we show that the full sample bootstrap method indeed works for the endpoint estimator proposed by Hall (1982). We organize this paper as follows. Section 2 gives the main results. A simulation study is given in Section 3.

## 2 Main results

Let  $\mathbf{X}_n^* = \{X_1^*, \dots, X_n^*\}$  be a bootstrap sample from the original sample  $\mathbf{X}_n = \{X_1, \dots, X_n\}$  and  $X_{n,1}^* \leq X_{n,2}^* \leq \dots \leq X_{n,n}^*$  be the corresponding order statistics. Then the bootstrap endpoint estimator is defined as the solution of the following equation

$$\frac{1}{k+1} \sum_{i=1}^k \frac{\theta - X_{n,n-k}^*}{\theta - X_{n,n-i+1}^*} \left\{ \frac{1}{k+1} \sum_{i=1}^k \log \frac{\theta - X_{n,n-i+1}^*}{\theta - X_{n,n-k}^*} + 1 \right\} = 1. \quad (5)$$

Let us denote this solution as  $\hat{\theta}_{Hall}^*(k)$ . Further we can define the bootstrap estimator for the index  $\gamma$  as

$$\hat{\gamma}_{Hall}^*(k) = \frac{1}{k+1} \sum_{i=1}^k \log \frac{\hat{\theta}_{Hall}^*(k) - X_{n,n-i+1}^*}{\hat{\theta}_{Hall}^*(k) - X_{n,n-k}^*}. \quad (6)$$

In order to derive the asymptotic distributions of these estimators, we need the following second order condition. Let  $U(t)$  denote the inverse function of  $1/(1-F(t))$ . Suppose that there exist functions  $a(t) > 0$  and  $A(t) \rightarrow 0$  such that

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \frac{1}{\rho} \left( \frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right) := H_{\gamma, \rho}(x) \quad (7)$$

for some  $\rho \leq 0$ . We refer to de Haan and Stadtmüller (1996) or de Haan and Ferreira (2006) for more details on the second order condition.

The following theorem shows that the bootstrap endpoint estimator is consistent.

**Theorem 2.1.** *Suppose (7) holds for some  $\gamma \in (-1/2, 0)$  and  $\rho < 0$ . Further assume that*

$$k = k(n) \rightarrow \infty \quad \text{and} \quad k^{1/2} A(n/k) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Then

$$\sup_{x \in R} |P(\hat{\theta}_{Hall}^*(k) - \hat{\theta}_{Hall}(k) \leq x | \mathbf{X}_n) - P(\hat{\theta}_{Hall}(k) - \theta \leq x)| \xrightarrow{P} 0 \quad (8)$$

and

$$\sup_{x \in R} |P(\sqrt{k} \frac{\hat{\theta}_{Hall}^*(k) - \hat{\theta}_{Hall}(k)}{\sigma^*(k)} \leq x | \mathbf{X}_n) - P(\sqrt{k} \frac{\hat{\theta}_{Hall}(k) - \theta}{\sigma(k)} \leq x)| \xrightarrow{P} 0 \quad (9)$$

as  $n \rightarrow \infty$ , where

$$\sigma(k) = X_{n,n-k} \left\{ \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n,n-i+1}}{X_{n,n-k}} \right\} \{1 - \hat{\gamma}_{Hall}(k)\} \{\hat{\gamma}_{Hall}(k)\}^{-2} \times \{1 + \hat{\gamma}_{Hall}(k)\} \{1 + 2\hat{\gamma}_{Hall}(k)\}^{1/2}$$

and

$$\sigma^*(k) = X_{n,n-k}^* \left\{ \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n,n-i+1}^*}{X_{n,n-k}^*} \right\} \{1 - \hat{\gamma}_{Hall}^*(k)\} \{\hat{\gamma}_{Hall}^*(k)\}^{-2} \times \\ \{1 + \hat{\gamma}_{Hall}^*(k)\} \{1 + 2\hat{\gamma}_{Hall}^*(k)\}^{1/2}.$$

*Remark 2.2.* As usual, the theoretical optimal choice of  $k$  is the one that minimizes the coverage error. However, it remains unknown on how to derive the expansion for the coverage probability. By noting that the condition of  $k^{1/2}A(n/k) \rightarrow 0$  implies that  $k = o(n^{-2\rho/(1-2\rho)})$ , one can use  $k$  as  $n^{-\delta\hat{\rho}/(1-2\hat{\rho})}$  for some  $\delta \in (0, 2)$  and estimator  $\hat{\rho}$  for  $\rho$ . In the simulation study, we employ  $\delta = 1.4, 1.5, 1.6, 1.7, 1.8$  and use the estimator  $\hat{\rho}$  defined in (2.2) of Fraga Alves et al. (2003).

### 3 Simulation study

First we investigate the finite sample behavior of the confidence interval constructed by bootstrapping the pivotal statistic  $T_{n,k} := \sqrt{k}\{\hat{\theta}_{Hall}(k) - \theta\}/\sigma(k)$  and compare it with the subsample bootstrap method based on statistic  $R_{2,n}$  defined in Section 4 of Athreya and Fukuchi (1997) in terms of both coverage probability and interval length.

We draw 1000 random samples of size  $n = 200, 1000, 2000$  from the random variable  $\theta - 1/Z$ , where  $Z$  has the *Burr*( $\tau_1, \tau_2$ ) distribution, i.e.,  $P(Z \leq z) = 1 - (1 + z^{\tau_1})^{-\tau_2}$  for  $z > 0$ . Note that (7) holds with  $\gamma = -1/(\tau_1\tau_2)$ ,  $\rho = -1/\tau_2$ ,  $a(t) = -\gamma t^\gamma - (\gamma + \rho)\tau_1^{-1}t^{\gamma+\rho}$  and  $A(t) = -\rho(\gamma + \rho)\tau_1^{-1}a^{-1}(t)t^{\gamma+\rho}$  for this considered distribution. Particularly, we consider  $(\tau_1, \tau_2) = (40, 1/10)$ ,  $(8, 1/2)$ , and  $\theta = 0$ . For calculating the subsample bootstrap intervals based on  $R_{2,n}$ , we choose  $m = n^{0.3}, n^{0.5}, n^{0.7}$ . We plot the empirical coverage probabilities for the above two confidence intervals with levels 0.9 and 0.95 against different values of  $k$ , medians of the interval lengths of these 1000 confidence intervals for each method.

Next, we calculate the coverage probability for the confidence interval based on  $T_{n,k}$  by choosing  $k$  as  $\hat{k}_1 = n^{-1.4\hat{\rho}/(1-2\hat{\rho})}$ ,  $\hat{k}_2 = n^{-1.5\hat{\rho}/(1-2\hat{\rho})}$ ,  $\hat{k}_3 = n^{-1.6\hat{\rho}/(1-2\hat{\rho})}$ ,  $\hat{k}_4 = n^{-1.7\hat{\rho}/(1-2\hat{\rho})}$ ,  $\hat{k}_5 = n^{-1.8\hat{\rho}/(1-2\hat{\rho})}$ ,  $k_1 = 50$ ,  $k_2 = 100$ ,  $k_3 = 150$ ,  $k_4 = 200$ , and  $k_5 = 250$ , where  $\hat{\rho}$  is defined in (2.2) of Fraga Alves et al. (2003). We consider  $n = 1000, 2000$ , and use the same setup as above. Coverage probabilities with levels 0.9 and 0.95 are reported in Tables 1 and 2, respectively. From these two tables, we observe that  $k = \hat{k}_4$  and  $\hat{k}_5$  work well, and coverage probability for the case of  $\tau_1 = 40$  is more accurate than that for the case of  $\tau_1 = 8$ .

In summary, the bootstrap method based on  $T_{n,k}$  for some range of  $k$  is comparable to the subsample bootstrap method based  $R_{2,n}$  in terms of coverage probability, but its median of interval lengths is smaller. Moreover, the proposed adaptive choice of  $k$  works well.

Table 1: Empirical coverage probabilities are reported for the confidence intervals with level 0.9 based on bootstrapping pivotal statistic with estimated  $k$  and fixed  $k$ .

$(n, \tau_1, \tau_2)$	(1000, 40, 1/10)	(1000, 8, 1/2)	(2000, 40, 1/10)	(2000, 8, 1/2)
$k = \hat{k}_1$	0.880	0.749	0.916	0.790
$k = \hat{k}_2$	0.898	0.818	0.922	0.838
$k = \hat{k}_3$	0.922	0.877	0.920	0.883
$k = \hat{k}_4$	0.919	0.911	0.897	0.906
$k = \hat{k}_5$	0.913	0.925	0.895	0.895
$k = k_1$	0.874	0.872	0.865	0.864
$k = k_2$	0.956	0.955	0.945	0.945
$k = k_3$	0.943	0.942	0.932	0.933
$k = k_4$	0.913	0.920	0.896	0.896
$k = k_5$	0.892	0.892	0.896	0.896

Table 2: Empirical coverage probabilities are reported for the confidence intervals with level 0.95 based on bootstrapping pivotal statistic with estimated  $k$  and fixed  $k$ .

$(n, \tau_1, \tau_2)$	(1000, 40, 1/10)	(1000, 8, 1/2)	(2000, 40, 1/10)	(2000, 8, 1/2)
$k = \hat{k}_1$	0.887	0.761	0.937	0.799
$k = \hat{k}_2$	0.909	0.825	0.948	0.843
$k = \hat{k}_3$	0.942	0.883	0.952	0.896
$k = \hat{k}_4$	0.946	0.921	0.953	0.930
$k = \hat{k}_5$	0.946	0.941	0.945	0.932
$k = k_1$	0.879	0.879	0.868	0.868
$k = k_2$	0.967	0.968	0.963	0.963
$k = k_3$	0.972	0.974	0.964	0.964
$k = k_4$	0.962	0.969	0.956	0.957
$k = k_5$	0.954	0.967	0.947	0.949

## References

- [1] Aarssen, K., and de Haan, L. (1994) "On the maximal life span of humans," *Mathematical Population Studies*, 4, 259–281.
- [2] Angus, J. (1993) "Asymptotic theory for bootstrapping the extremes," *Communications in Statistics - Theory and Methods*, 22, 15–30.
- [3] Athreya, K.B. and Fukuchi, J.I. (1997) "Confidence intervals for endpoints of a c.d.f. via bootstrap," *Journal of Statistical Planning and Inference*, 58, 299–320.
- [4] Bickel, P.J. and Freedman, D.A. (1981) "Some asymptotic theory for the bootstrap," *Annals of Statistics*, 9, 1196–1217.
- [5] De Haan, L. and Ferreira, A. (2006) *Extreme Value Theory-An Introduction*, Springer.
- [6] De Haan, L. and Stadtmüller, U. (1996) "Generalized regular variation of second order," *Journal of the Australian Mathematical Society*, A, 61, 381–395.
- [7] Deheuvels, P., Mason, D. and Shorack, G. (1993) "Some results on the influence of extremes on the bootstrap," *Annales de l'Institut Henri Poincaré*, 29, 83–103.

- [8] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989) "A moment estimator for the index of an extreme-value distribution," *Annals of Statistics*, 17, 1833–1855.
- [9] Drees, H., Ferreira, A. and de Haan, L. (2004) "On maximum likelihood estimation of the extreme value index," *Annals of Applied Probability*, 14, 1179–1201.
- [10] Ferreira, A., de Haan, L. and Peng, L. (2003) "On optimising the estimation of high quantiles of a probability distribution," *Statistics*, 37, 403–434.
- [11] Fraga Alves, M.I., de Haan, L. and Lin, T. (2003) "Estimation of the parameter controlling the speed of convergence in extreme value theory," *Mathematical Methods of Statistics*, 12, 155–176.
- [12] Geluk, J. and de Haan, L. (2002) "On bootstrap sample size in extreme value theory," *Publ. Inst. Math. (Beograd)(N.S.)*, 71, 21–25.
- [13] Hall, P. (1982) "On estimating the endpoint of a distribution," *Annals of Statistics*, 10, 556–568.
- [14] Hall, P. and Wang, J.Z. (1999) "Estimating the end-point of a probability distribution using minimum-distance methods," *Bernoulli*, 5, 177–189.
- [15] Li, D. and Peng, L. (2009) "Does Bias Reduction with External Estimator of Second Order Parameter Work for endpoint?" *Journal of Statistical Planning and Inference*, 139, 1937–1952.
- [16] Li, D. and Peng, L. (2010) "Compare extreme models when the sign of the extreme value index is known," *Statistics & Probability Letters*, 80, 739–746.
- [17] Loh, W.Y. (1984) "Estimating an endpoint of a distribution with resampling methods," *Annals of Statistics*, 12, 1534–1550.
- [18] Peng, L. and Qi, Y. (2009) "Maximum likelihood estimation of extreme value index for irregular cases," *Journal of Statistical Planning and Inference*, 139, 3361–3376.
- [19] Qi, Y. (2008) "Bootstrap and empirical likelihood methods in extremes," *Extremes*, 11, 81–97.
- [20] Smith, R.L. (1987) "Estimating tails of probability distributions," *Annals of Statistics*, 15, 1174–1207.
- [21] Swanepoel, J.W.H. (1986) "A note on proving that the (modified) bootstrap works," *Communications in Statistics - Theory and Methods*, 15, 3193–3203.
- [22] Woodroffe, M. (1974) "Maximum likelihood estimation of translation parameter of truncated distribution II," *Annals of Statistics*, 2, 474–488.