# Confidence Intervals for the Tail Index of the Pareto Distribution of the First Type

by

BO YUN MIN

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Statistics
University of Manitoba
Winnipeg

**Abstract**

Using pivotal quantities, we construct a variety of exact and asymptotic confidence intervals (CI) for the tail index (shape parameter) $\alpha > 0$ of the Pareto distribution of the first type, Pareto(I)($\alpha$,$\sigma$), assuming that the scale parameter $\sigma > 0$ is known. The obtained CI's are compared in terms of their expected lengths and finite-sample coverage probabilities, and thus the better performing CI's among them are determined. We also outline the construction of exact and asymptotic CI's for $\alpha$ when $\sigma$ is unknown.

# Acknowledgments

First of all, I would like to give my thanks to my supervisor Dr. Yuliya Martsynyuk without whom I could have never carried out my work properly and who was not only always available for me, but was also extremely kind in providing guidance and comments regarding my thesis.

Further, I would like to thank my examining committee members Dr. Elif Acar and Dr. Felicia Magpantay for their insightful inputs and suggestions.

I would also want to thank Mr. Dave Gabrielson, the Information Technologist of the Department of Statistics, who created the Computational Cluster for the Department of Statistics which I could not do without all through my work, and who also taught me how to use LaTeX.

My gratitude also goes to all the faculty members of the Department of Statistics who taught me invaluable knowledge: Dr. Andrew Morris, Dr. Zenaida Mateo, Dr. Saumen Mandal, Dr. Alexandre Leblanc, Dr. Liqun Wang, Dr. Saman Muthukumarana, Dr. Po Yang, Dr. Brad Johnson, Dr. Elif Acar, and Dr. Mohammad Jafari Jozani.

I greatly appreciate the generous fundings I received from my supervisor, from the Department of Statistics and the Faculty of Science, and from the GETS program of the Faculty of Graduate Studies.

Last but not least, I am grateful to my family and friends who have endlessly been supportive and kind.

# Contents

# List of Tables

# Notations and Abbreviations

| | |
|---|---|
| r.v. | Random variable |
| i.i.d. | Independent and identically distributed |
| Pareto(I)($\alpha$,$\sigma$) | Pareto distribution (or r.v.) of the first type with tail index (shape parameter) $\alpha$ and scale parameter $\sigma$ |
| $\chi^2_n$ | Chi-square distribution (or r.v.) with $n$ degrees of freedom |
| $Exp(\alpha)$ | Exponential distribution (or r.v.) with mean $1/\alpha$ |
| $\Gamma(\alpha, \beta)$ | Gamma distribution (or r.v.) with mean $\alpha/\beta$ |
| $N(\mu, \sigma^2)$ | Normal distribution (or r.v.) with mean $\mu$ and variance $\sigma^2$ |
| $IG(\alpha, \beta)$ | Inverse gamma distribution (or r.v.) with shape parameter $\alpha$ and rate parameter $\beta$ |
| := | Equality by definition |

| | |
|---|---|
| $\sim$ | Equality in distribution of two random variables |
| $\Leftrightarrow$ | If and only if |
| $\mathbb{1}_{\{\cdot\}}$ | Indicator function of an event |
| $E(X)$ | Expected value of random variable $X$ |
| $E(X^v)$ | Moment of order $v$ of random variable $X$ |
| $E\|X\|^v$ | Absolute moment of order $v$ of random variable $X$ |
| $Var(X)$ | Variance of random variable $X$ |
| $D$[0,1] | The space of real-valued functions on [0,1] that are right-continuous and have left-hand limits |
| $\{W(t), 0 \leq t \leq 1\}$ | Standard Wiener process on [0,1] |
| $z_{\alpha/2}$ | $1 - \alpha/2$ quantile of the standard normal distirbution |
| $X_{1:n}$ | The first order statistic of random sample $X_1$,...,$X_n$ |
| $[x]$ | Floor function of $x$; the greatest integer less than or equal to $x$ |
| $\xrightarrow{P}$ | Convergence in probability |
| $op(1)$ | Sequence of r.v.'s that converges in probability to 0 |
| $\xrightarrow{D}$ | Convergence in distribution |
| CI | Confidence interval |

| | |
|---|---|
| PDF | Probability density function |
| DAN | Domain of attraction of the normal law |
| ARE | Asymptotic relative efficiency |
| MLE | Maximum likelihood estimator |
| MME | Method of moments estimator |
| GME | Generalized median estimator |
| CLT | Central limit theorem |
| FCLT | Functional central limit theorem |
| WLLN | Weak law of large numbers |

# Definitions

$$\bar{X}_n := \frac{\sum_{i=1}^{n} X_i}{n}$$

$$S_n(X) := \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n-1}}$$

$$T_n(X) := \frac{\sum_{i=1}^{n} X_i}{S_n(X)\sqrt{n}}$$

$$T_n^t(X) := \frac{\sum_{i=1}^{[nt]} X_i}{S_n(X)\sqrt{n}}, \qquad \text{for } 0 \leq t \leq 1$$

$$\widehat{CP}_{AI_k} := \frac{\sum_{i=1}^{10{,}000} \mathbb{1}_{\{\alpha\, \in AI_k \text{ for the } i^{\text{th}} \text{ sample}\}}}{10,000}$$

(Empirical coverage probability of confidence interval $AI_k$)

$$\widehat{E(\text{Length of } AI_k)} := \frac{\sum_{i=1}^{10000}(\text{Length of } AI_k \text{ based on the } i^{\text{th}} \text{ sample})}{10,000}$$

(Empirical expected length of confidence interval $AI_k$)

$$\hat{r}_k := \frac{\widehat{E(\text{Length of } AI_k)}}{\widehat{E(\text{Length of } AI_1)}}$$

(Ratio of empirical expected lengths of confidence intervals $AI_k$ and $AI_1$)

# Chapter 1

# Introduction

The Pareto distribution of the first type, denoted by Pareto(I)($\alpha$,$\sigma$), has the probability density function

$$f(x) = \begin{cases} \dfrac{\alpha\sigma^{\alpha}}{x^{\alpha+1}} & , \quad \text{if } x \geq \sigma, \\ 0 & , \quad \text{if } x < \sigma, \end{cases} \tag{1.1}$$

where $\alpha > 0$ is the tail index (shape parameter) and $\sigma > 0$ is the scale parameter of the distribution. The parameter $\alpha$ shows how heavy the tail of the distribution is, and $E(X^v) < \infty$ for $0 < v < \alpha$. This distribution was named after Vilfredo Pareto, who in 1897 made a famous claim that he could model the proportion of people in a society whose income exceeded $x$ as $Cx^{-\alpha}$, with some positive real constants $C$ and $\alpha$ (cf. Arnold (1983)).

Pareto(I)($\alpha$,$\sigma$) is a part of a large family of univariate and multivariate Pareto distributions (cf. Arnold (1983) and Johnson, Kotz, and Balakrishnan

(1994, Chapter 20)). These distributions have been widely applied in actuarial science, econometrics, and other fields. In particular, they have been useful for modelling income and wealth distributions and insurance claims. However, while Pareto distributions are good at approximating the distributions of the upper values of many variables, the same cannot be said for lower values. When modelling income distributions, for example, Pareto distributions are most useful for examining the wealthiest members of a society. Also, in modelling of insurance claims, Pareto distributions are useful for modelling large losses above a certain threshold. For examples of some real-life data sets that are described with Pareto distributions, we refer to Brazauskas and Serfling (2003) and Arnold (1983).

This thesis focuses on confidence intervals (CI) for the tail index $\alpha$ of Pareto(I)($\alpha$,$\sigma$) mainly under the assumption that the scale parameter $\sigma$ is known. In applications, estimation of $\alpha$ is often of more interest than that of $\sigma$ and, moreover, in some situations, it is indeed reasonable to assume that $\sigma$ is known. For example, when modelling insurance claims with the Pareto(I)($\alpha$,$\sigma$) distribution, $\sigma$ may represent a deductible that is set in advance. Furthermore, dealing with robust and efficient point estimation of $\alpha$ when $\sigma$ is assumed to be known, Brazauskas and Serfling (2000a, section 6) argue that even small improvements in methods of estimation of $\alpha$ can yield significant favourable impact in estimating important quantities based on $\alpha$ in applications. This inspired us to closely investigate precision of the CI's for $\alpha$ that we construct in the thesis. Accordingly, our CI's are compared from the standpoint of two standard criteria, expected length and finite-sample coverage probability, and thus we determine the better performing CI's, that is those with relatively short expected lengths and appropriately high finite-sample coverage probabilities.

While many of the pivotal and asymptotically pivotal quantities we use for building our CI's for $\alpha$ have been well known, we have not found any references with comparative analysis of properties of such CI's. Moreover, the literature on CI's for $\alpha$ constructed with other possible pivots is scarce. On the other hand, the research in this thesis is largely inspired by Martsynyuk and Tuzov (2016) and the earlier mentioned Brazauskas and Serfling (2000a). Martsynyuk and Tuzov (2016) constructed CI's for a mean of a population from the domain of attraction of the normal law that are based on convergence in distribution of five special functionals of the so-called Student process, that is on five special cases of the functional central limit theorem (FCLT). They concluded that the obtained CI's may be preferred to a classical asymptotic CI that follows simply from asymptotic normality of the Student $t$-statistic. Inspired by Martsynyuk and Tuzov (2016), we adapt convergence in distribution of a few of their functionals to constructing asymptotic CI's for $\alpha$ of Pareto(I)($\alpha$,$\sigma$), assuming that $\sigma$ is known (cf. section 3.2). This way of constructing asymptotic CI's for $\alpha$ appears to be new. Assuming that $\sigma$ is known, Brazauskas and Serfling (2000a) studied asymptotic performance of various point estimators of $\alpha$ of Pareto(I)($\alpha$,$\sigma$) based on two criteria: asymptotic relative efficiency (with respect to the maximum likelihood estimator (MLE) of $\alpha$) and robustness against upper outliers (cf. subsection 3.3.3). In particular, in their search for a relatively highly efficient and adequately robust estimator of $\alpha$, Brazauskas and Serfling (2000a) introduced the so-called generalized median estimator (GME) and showed that it performs best in this sense among several other estimators of $\alpha$. Based on asymptotic normality of this estimator, we construct one of our best performing asymptotic CI's for $\alpha$ in this thesis, denoted by $AI_{10}$ (cf. subsection 3.3.3 and Chapter 4 for details).

In Chapter 2, using known exact pivots that are based on the MLE's for $\alpha$ and $\sigma$, we derive two exact CI's for $\alpha$ assuming that $\sigma$ is known and investigate their expected lengths (cf. section 2.1). We minimize the expected length of a shorter CI numerically using the R 3.3.2 software and writing an algorithm for finding the appropriate lower and upper quantiles of a chi-square distribution. An analogue of this interval in the case of $\sigma$ being unknown is presented in section 2.2. All R codes of this thesis are collected in Appendix.

In Chapter 3, we assume that $\sigma$ is known and build a variety of asymptotic CI's for $\alpha$ from several asymptotically pivotal quantities. Accordingly, sections 3.1 and 3.2 utilize respectively the central limit theorem (CLT), with asymptotic normality of the Student $t$-statistic, and convergence in distribution of special functionals of the corresponding Student process, as special cases of the FCLT for the latter process. While the use of such CLT and FCLT requires that we have a distribution from the domain of attraction of the normal law (DAN) and hence imposes the condition $\alpha \geq 2$ on Pareto(I)($\alpha$,$\sigma$), we avoid any restrictions on $\alpha$ by simply converting Pareto(I)($\alpha$,$\sigma$) to an exponential distribution with mean $1/\alpha$, $Exp(\alpha)$ (that belongs to DAN for any $\alpha > 0$), and by applying the CLT and FCLT for the latter distribution. Studying Pareto(I)($\alpha$,$\sigma$) with $Exp(\alpha)$ has been a standard practice in the literature. Moreover, according to our simulation studies, this approach also leads to the better performing CI's for $\alpha$. In section 3.3, we build CI's for $\alpha$ using asymptotic normality of three estimators of $\alpha$: the MLE, method of moments estimator, and GME of Brazauskas and Serfling (2000a). In sum, we end up with studying ten asymptotic CI's for $\alpha$, $AI_1$ to $AI_{10}$, in sections 3.1 - 3.3. For a convenient comparison of all these CI's, each $AI_k$ of $AI_2$ to $AI_{10}$ is compared to the CLT based $AI_1$ of (3.18) via studying the finite-sample coverage probabilities of $AI_k$ and $AI_1$ and the ratio of the expected lengths of $AI_k$ to that of $AI_1$. The comparisons are done numerically through simulations in R 3.3.2, since it is not feasible to obtain closed-form expressions for most of the expected lengths and finite-sample coverage probabilities of $AI_1$ - $AI_{10}$ (cf. Tables 3.2 - 3.10). The last section of Chapter 3, section 3.4, briefly discusses some possible adaptations of the methods in sections 3.1 - 3.3 to construction of CI's for $\alpha$ when $\sigma$ is unknown.

Finally, in Chapter 4, we present a summary table of the performance of the asymptotic CI's $AI_2$ - $AI_{10}$ versus that of $AI_1$ of (3.18), Table 4.1, that is based on Tables 3.2 - 3.10 produced in Chapter 3. This enables us to determine the better performing asymptotic CI's for $\alpha$. On one hand, most of $AI_2$ - $AI_{10}$ demonstrate some trade-off between an expected length and finite-sample coverage probability of a CI, and are a bit longer than $AI_1$ on average, but have somewhat higher finite-sample coverage probabilities than those of $AI_1$. However, the CI $AI_7$ of (3.51), based on the asymptotic normality of the MLE of $\alpha$, performs better than $AI_1$ in terms of both criteria. On the other hand, while the expected length of $AI_{10}$ of (3.63), the CI built from the asymptotic normality of the GME of $\alpha$, is under, or slightly above, that of $AI_1$, the finite-sample coverage probabilities of $AI_{10}$ are the highest among $AI_1$ - $AI_{10}$. Consequently, $AI_{10}$ may be desirable when having a CI for $\alpha$ with a higher finite-sample coverage probability is a priority. We also conclude that the performance of the asymptotic CI's $AI_7$ and $AI_{10}$ is very comparable to that of the exact CI $I_1$ of (2.7) in terms of the finite-sample coverage probabilities and expected lengths, where $I_1$ is the better performing of the two exact CI's for $\alpha$ from section 2.1. However, we believe that in the case when the actual data departures from the assumed Pareto(I)($\alpha$,$\sigma$) distribution, for example due to unrepresentative outliers in the sample, $AI_{10}$ likely outperforms $I_1$, and, perhaps, this is also true for $AI_7$ and some other asymptotic CI's for $\alpha$ in the thesis. Further to this conjecture, we conclude Chapter 4 with a list of other possible extensions of the research of this thesis.

# Chapter 2

# Exact Confidence Intervals for the Tail Index $\alpha$

In this chapter, we construct confidence intervals for the tail index $\alpha$ of the Pareto(I)($\alpha$,$\sigma$) distribution that are based on exact pivotal quantities. Expected lengths of the obtained intervals are also investigated. In section 2.1, we consider the case when the scale parameter $\sigma$ is known, while section 2.2 addresses the situation of $\sigma$ being unknown. We minimize numerically the expected length of the shorter of the two CI's built in section 2.1 and also that of the CI of section 2.2.

## 2.1 Scale parameter $\sigma$ is known

Let $\chi_v^2$ denote a chi-square distribution or random variable (r.v.) with $v$ degrees of freedom, $v$=1,2,..., and $\sim$ stands for the equality in distribution of two r.v.'s.

We first consider the following pivotal quantity (cf. Arnold (1983, p. 225), for example):

$$\frac{2n\alpha}{\hat{\alpha}_{mle}} \sim \chi^2_{2n}, \tag{2.1}$$

where

$$\hat{\alpha}_{mle} = \frac{n}{\sum_{i=1}^{n} log(\frac{X_i}{\sigma})} \tag{2.2}$$

is the maximum likelihood estimator (MLE) of the tail index $\alpha$ derived under the assumption that $\sigma$ is known. The relationship in (2.1) can be verified by using the transformations

$$log\Big(\frac{X_i}{\sigma}\Big) \sim Exp(\alpha) \quad \text{and} \quad \sum_{i=1}^{n} log\Big(\frac{X_i}{\sigma}\Big) \sim \Gamma(n, \alpha), \tag{2.3}$$

where $Exp(\alpha)$ and $\Gamma(n, \alpha)$ stand for an exponential and gamma distributions (or r.v.'s) with the means $1/\alpha$ and $n/\alpha$, respectively, as well as by using that

$$2\alpha\Gamma(n, \alpha) \sim \Gamma\left(n, \frac{1}{2}\right) \quad \text{and} \quad \Gamma\left(n, \frac{1}{2}\right) \sim \chi^2_{2n}. \tag{2.4}$$

For $0 < \gamma < 1$, let

$$P(\chi^2_{2n} < \chi^2_{2n,\gamma_1}) = \gamma_1 \quad \text{and} \quad P(\chi^2_{2n} > \chi^2_{2n,1-\gamma_2}) = \gamma_2, \tag{2.5}$$

where $\gamma_1$+$\gamma_2$=$\gamma$. Then, from (2.1) and (2.5),

$$P\Big(\chi^2_{2n,\gamma_1} \leq \frac{2n\alpha}{\hat{\alpha}_{mle}} \leq \chi^2_{2n,1-\gamma_2}\Big) = 1 - \gamma, \tag{2.6}$$

which yields the following $100(1-\gamma)\%$ exact CI for $\alpha$:

$$I_1 := \left[ \frac{\chi^2_{2n,\gamma_1}}{2\sum_{i=1}^n log(\frac{X_i}{\sigma})}, \frac{\chi^2_{2n,1-\gamma_2}}{2\sum_{i=1}^n log(\frac{X_i}{\sigma})} \right]. \tag{2.7}$$

To find the expected length of $I_1$, we first compute $E\left[1/\sum_{i=1}^n log(X_i/\sigma)\right]$ using (2.3).

$$\begin{aligned} E\left[\frac{1}{\sum_{i=1}^n log(\frac{X_i}{\sigma})}\right] &= \int_0^\infty \frac{1}{y}\,\frac{\alpha^n}{\Gamma(n)}\,y^{n-1}\,e^{-\alpha y}\,dy \\ &= \frac{\alpha^n}{\Gamma(n)}\,\frac{\Gamma(n-1)}{\alpha^{n-1}}\int_0^\infty \frac{\alpha^{n-1}}{\Gamma(n-1)}\,y^{(n-1)-1}\,e^{-\alpha y}\,dy \\ &= \frac{\alpha^n}{\Gamma(n)}\,\frac{\Gamma(n-1)}{\alpha^{n-1}} \\ &= \frac{\alpha}{n-1}. \end{aligned} \tag{2.8}$$

Thus, the expected length of $I_1$ is

$$E(\text{Length of } I_1) = \frac{\alpha}{2(n-1)}(\chi^2_{2n,1-\gamma_2} - \chi^2_{2n,\gamma_1}). \tag{2.9}$$

It is minimized when $\chi^2_{2n,1-\gamma_2} - \chi^2_{2n,\gamma_1}$ is, and in order to find such a lower and upper quantile of the $\chi^2_{2n}$ distribution, denoted by $a$ and $b$ for easier reference, we apply Theorem 9.3.2 in Casella and Berger (2002). Accordingly, $a$ and $b$

must satisfy the following conditions:

$$\begin{aligned}
&(i)\ \int_a^b g(x)\,dx \ = \ 1-\gamma, \\
&(ii)\ g(a) \ = \ g(b) \ > 0,\ \text{and} \\
&(iii)\ a \le x^* \le b,\ \text{where } x^* \text{ is the mode of } g(x),
\end{aligned} \tag{2.10}$$

where $g(\cdot)$ denotes the probability density function (PDF) of our unimodal $\chi^2_{2n}$ distribution.

Next, we find $a$ and $b$ as in (2.10) numerically using the R 3.3.2 software (all our R codes are included in Appendix). In fact, having taken $a \in (0, x^*)$, we find $b > x^*$ from the condition $(ii)$ in (2.10) using the *uniroot* function in R, and then we check if $a$ and $b$ in hand satisfy $(i)$ in (2.10). To optimize our search, we apply the following logic. We first take $a_1 = 0.45x^*$ and $a_2 = 0.9999x^*$ and find the corresponding $b_1$ and $b_2$. For all $n = 50$, 100, 300, and 1000 and $\gamma = 0.02$, 0.05, and 0.1 considered in these computations, we have $\int_{a_1}^{b_1} g(x)dx > 1 - \gamma + 10^{-7}$ and $\int_{a_2}^{b_2} g(x)dx < 1 - \gamma - 10^{-7}$. Consequently, the desired lower quantile must be between $a_1$ and $a_2$. On the next step, we take $a_3 = (a_1 + a_2)/2$ and find the corresponding $b_3$. If $\int_{a_3}^{b_3} g(x)dx$ is within our tolerance level $10^{-7}$ of $1 - \gamma$, then we stop our search and declare $a_3$ and $b_3$ our desired pair of quantiles. Otherwise, we continue our search by considering the next candidate $a_4 = (a_1 + a_3)/2$, if $\int_{a_3}^{b_3} g(x)dx < 1 - \gamma - 10^{-7}$, or $a_4 = (a_2 + a_3)/2$, if $\int_{a_3}^{b_3} g(x)dx > 1 - \gamma + 10^{-7}$. We repeat this process until we find an interval [a lower quantile, an upper quantile] over which $\int g(x)dx$ is within $10^{-7}$ of $1-\gamma$. As a result of this algorithm, we obtain Table 2.1 of the lower and upper quantiles $a$ and $b$ satisfying (2.10).

| | $1-\gamma$ | | |
|---|---|---|---|
| $n$ | 0.9 | 0.95 | 0.98 |
| 50 | (76.706,122.911) | (73.021,128.114) | (68.891,134.340) |
| 100 | (167.022,232.591) | (161.486,239.642) | (155.209,248.015) |
| 300 | (542.890,656.719) | (532.737,668.387) | (521.094,682.125) |
| 1000 | (1895.810,2103.798) | (1876.641,2124.482) | (1854.516,2148.700) |

Table 2.1: Quantiles ($a$,$b$) of $\chi^2_{2n}$ as in (2.10)

We note in passing that the way of finding $a$ and $b$ described above turns out to be more time efficient than an alternative one when we first take a large number of quantile pairs $a$ and $b$ satisfying $(i)$ in (2.10) (based on a grid of $(0, x^*)$) and then select a pair among them for which $|g(a)-g(b)|$ is the smallest (cf. $(ii)$ of (2.10)).

Although in this section we assume that $\sigma$ is known and does not need to be estimated, let's consider the MLE $\tilde{\sigma}_{mle}$ for $\sigma$,

$$\tilde{\sigma}_{mle} = X_{1:n}, \tag{2.11}$$

which is commonly used when both $\alpha$ and $\sigma$ are unknown (cf. Arnold (1983)), where $X_{1:n} = min(X_1, ..., X_n)$. Using (2.3) and a known fact that

$$\tilde{\sigma}_{mle} \sim \text{Pareto(I)}(n\alpha, \sigma) \tag{2.12}$$

(cf. Arnold (1983, p. 225)), we have

$$n\alpha log\Big(\frac{X_{1:n}}{\sigma}\Big) \sim Exp(1). \tag{2.13}$$

The pivotal quantity in (2.13) can be used to construct another $(1-\gamma)100\%$ exact CI for $\alpha$ when $\sigma$ is assumed to be known:

$$I_2 := \left[\frac{a}{nlog(\frac{X_{1:n}}{\sigma})}, \frac{b}{nlog(\frac{X_{1:n}}{\sigma})}\right], \tag{2.14}$$

where positive $a$ and $b$ are such that $P(\alpha < Exp(1) < b) = 1-\gamma$, $0<\gamma<1$.

However, the expected length of $I_2$ is $\infty$, since

$$\begin{aligned}
E\left[\frac{1}{log(\frac{X_{1:n}}{\sigma})}\right] &= \int_0^\infty \frac{1}{u} n\alpha\, e^{-n\alpha u}\, du \\
&\geq \int_0^1 \frac{1}{u} n\alpha\, e^{-n\alpha u}\, du \geq \int_0^1 \frac{1}{u} n\alpha\, e^{-n\alpha}\, du \\
&= n\alpha e^{-n\alpha} \int_0^1 \frac{1}{u}\, du = \infty.
\end{aligned}$$

This means that in practice $I_2$ may have a very large average length and is not very desirable.

Based on the expected lengths of $I_1$ of (2.7) and $I_2$ of (2.14), we conclude that $I_1$ should be used when an exact CI for $\alpha$ is sought.

## 2.2 Scale parameter $\sigma$ is unknown

The two CI's of section 2.1 were obtained based on the assumption that $\sigma$ is known. Now, let's assume that $\sigma$ is unknown and build a CI for $\alpha$.

From Chapter 5 of Arnold (1983), the MLE's $\tilde{\sigma}_{mle}$ of $\sigma$ and $\tilde{\alpha}_{mle}$ of $\alpha$ are respectively (2.11) and

$$\tilde{\alpha}_{mle} = \frac{n}{\sum_{i=1}^{n} log\left(\frac{X_i}{\tilde{\sigma}_{mle}}\right)}, \tag{2.15}$$

where

$$\tilde{\alpha}_{mle} \sim IG(n-1, \alpha n), \tag{2.16}$$

that is $\tilde{\alpha}_{mle}$ follows the inverse gamma distribution with the shape parameter $n-1$ and the rate parameter $\alpha n$, and the PDF

$$f_{\tilde{\alpha}_{mle}}(x) = \frac{(\alpha n)^{n-1}}{\Gamma(n-1)x^n} e^{-\frac{n\alpha}{x}}, \quad x > 0,$$

where $\Gamma(\cdot)$ is the gamma function.

Now, we consider the following pivotal quantity that is based on $\tilde{\alpha}_{mle}$ (cf. Arnold (1983, p. 225)):

$$\frac{2\alpha n}{\tilde{\alpha}_{mle}} \sim \chi^2_{2n-2}. \tag{2.17}$$

The relationship (2.17) can be easily seen by first showing that (2.16) implies that

$$\frac{1}{\tilde{\alpha}_{mle}} \sim \Gamma(n-1, \alpha n) \tag{2.18}$$

and then applying (2.4). Using (2.17), we construct a $100(1-\gamma)\%$ exact CI for $\alpha$:

$$\left[ \frac{\chi^2_{2n-2,a}}{2\sum_{i=1}^{n} log\left(\frac{X_i}{\tilde{\sigma}_{mle}}\right)}, \frac{\chi^2_{2n-2,b}}{2\sum_{i=1}^{n} log\left(\frac{X_i}{\tilde{\sigma}_{mle}}\right)} \right], \tag{2.19}$$

where $\chi^2_{2n-2,a}$ and $\chi^2_{2n-2,b}$ are positive quantiles of $\chi^2_{2n-2}$ satisfying

$$P(\chi^2_{2n-2,a} \leq \chi^2_{2n-2} \leq \chi^2_{2n-2,b}) = 1-\gamma. \tag{2.20}$$

We notice that the CI in (2.19) is similar in form to that in (2.7) obtained under the assumption that $\sigma$ is known. Likewise to Table 2.1, the following Table 2.2 provides the quantiles $(\chi^2_{2n-2,a}, \chi^2_{2n-2,b})$ that minimize the expected length of (2.19).

| | $1-\gamma$ | | |
|---|---|---|---|
| $n$ | 0.9 | 0.95 | 0.98 |
| 50 | (74.942,120.676) | (71.302,125.833) | (67.224,132.008) |
| 100 | (165.187,230.425) | (159.684,237.445) | (153.443,245.781) |
| 300 | (540.985,654.624) | (530.850,666.274) | (519.228,679.990) |
| 1000 | (1893.862,2101.746) | (1874.703,2122.420) | (1852.590,2146.627) |

Table 2.2: Quantiles $(\chi^2_{2n-2,a}, \chi^2_{2n-2,b})$ of $\chi^2_{2n-2}$ minimizing the length of (2.19)

# Chapter 3

# Asymptotic Confidence Intervals for the Tail Index $\alpha$

In this chapter, we construct a variety of asymptotic CI's for $\alpha$ that are based on convergence in distribution of several asymptotically pivotal quantities.

We start off with the central limit theorem (CLT) for the Student $t$-statistic in section 3.1. Then, in section 3.2, we utilize special cases of the so-called functional CLT (FCLT) for the corresponding Student process. We discover that the CLT and FCLT in sections 3.1 and 3.2 based on the exponential transformation in (2.3) of the original Pareto(I)($\alpha$,$\sigma$) sample not only allow us to drop the condition $\alpha \geq 2$ required when they are applied for the Pareto(I)($\alpha$,$\sigma$) sample, but also lead to the better performing CI's for $\alpha$. Section 3.3 deals with the asymptotic normality of the MLE, method of moments estimator and generalized median estimator for $\alpha$. In sections 3.1 - 3.3, we assume throughout that the scale parameter $\sigma$ is known.

We also investigate numerically the expected lengths and finite-sample coverage probabilites of the asymptotic CI's obtained in sections 3.1 - 3.3, denoted by $AI_1$ to $AI_{10}$ (it is not feasible to obtain closed-form expressions for most of the expected lengths and finite-sample coverage probabilities of $AI_1$-$AI_{10}$). We do so by using 10,000 repetitions of a Pareto(I)($\alpha$,$\sigma$) sample of size $n$ and computing the empirical finite-sample probabilities and expected lengths of $AI_k$'s by the formulae:

$$\widehat{CP}_{AI_k} := \frac{\sum_{i=1}^{10{,}000} \mathbb{1}_{\{\alpha\ \in AI_k \text{ for the } i^{\text{th}} \text{ sample}\}}}{10{,}000}, \tag{3.1}$$

with an indicator function $\mathbb{1}_{\{\cdot\}}$ of an event, and

$$\widehat{E(\text{Length of } AI_k)} := \frac{\sum_{i=1}^{10000}(\text{Length of } AI_k \text{ based on the } i^{\text{th}} \text{ sample})}{10{,}000}. \tag{3.2}$$

In fact, each $AI_k$ of $AI_2$ to $AI_{10}$ is compared numerically to the CLT based $AI_1$ of (3.18) via studying the empirical finite-sample coverage probabilities of $AI_k$ and $AI_1$ and the ratio

$$\hat{r}_k := \frac{\widehat{E(\text{Length of } AI_k)}}{\widehat{E(\text{Length of } AI_1)}} \tag{3.3}$$

of their empirical expected lengths. Accordingly, Tables 3.2 - 3.10 are produced for the sample size $n$ = 50, 100, 300, and 1,000, the confidence level $1-\gamma$ = 0.9, 0.95 and 0.98, the tail index $\alpha$ = 0.5, 1.5, 2, 3, and 5 (representing various spread of Pareto(I)($\alpha$,$\sigma$), from big to small, but not too small so that the distribution would not behave like a degenerate one), and for $\sigma$=1 (we believe that the corresponding simulation results for other values of $\sigma$ should be similar). The entries in Tables 3.2 - 3.10 are in the form of

$$\hat{r}_k\,(\widehat{CP}_{AI_k}, \widehat{CP}_{AI_1}). \tag{3.4}$$

Finally, in section 3.4, we briefly discuss some possible adaptations of the methods in sections 3.1 - 3.3 to construction of CI's for $\alpha$ when the parameter $\sigma$ is unknown.

## 3.1 Student $t$-statistic based confidence intervals

First of all, let's look at the notion of the domain of attraction of the normal law (DAN). Let $X, X_1, X_2, ...$ be independent and identically distributed (i.i.d.) random variables (r.v.'s). Then, $X$ is said to belong to DAN if there exist normalizing sequences $a_n$ and $b_n$ such that

$$\frac{\sum_{i=1}^{n} X_i - b_n}{a_n} \xrightarrow{D} N(0,1), \ \ n \to \infty, \tag{3.5}$$

(cf. Gut (2009), for example).

Note that the classical CLT (under $0 < Var(X) < \infty$) is a special case of (3.5): $b_n$ and $a_n$ can be taken as $nE(X)$ and $\sqrt{nVar(X)}$, respectively. In fact, (3.5) is less restrictive than the CLT, as it is known that if $X \in$ DAN, then $E|X|^v < \infty$ for all $v \in (0,2)$, while $Var(X)$ does not need to be finite.

Does the Pareto(I)($\alpha$,$\sigma$) distribution belong to DAN? It can be easily seen that for $\alpha > 2$, the second moment of Pareto(I)($\alpha$,$\sigma$) is finite and hence, if $X \in$ Pareto(I)($\alpha$,$\sigma$), then the CLT holds for $X_1, X_2, \ldots$. Thus, Pareto(I)($\alpha$,$\sigma$) belongs to DAN for $\alpha > 2$. When $\alpha < 2$, Pareto(I)($\alpha$,$\sigma$) is not in DAN, as in this case its $\alpha^{\text{th}}$ moment is infinite. In order to see if Pareto(I)(2,$\sigma$) $\in$ DAN, we can use the following useful characterization of Lévy (1937):

$$X \in \text{DAN} \quad \Leftrightarrow \quad \lim_{x\to\infty} \frac{x^2 P(|X| > x)}{E(X^2 \mathbb{1}_{|X|\le x})} = 0. \tag{3.6}$$

Let $X \sim$ Pareto(I)(2,$\sigma$), then

$$\lim_{x\to\infty} \frac{x^2 P(|X| > x)}{E(X^2 \mathbb{1}_{|X|\le x})} = \lim_{x\to\infty} \frac{x^2 \int_x^\infty \frac{2\sigma^2}{y^3} dy}{\int_\sigma^x y^2 \frac{2\sigma^2}{y^3} dy}$$

$$= \lim_{x\to\infty} \frac{x^2 \cdot (-y^{-2})\Big|_x^\infty}{log\, y\Big|_\sigma^x} = \lim_{x\to\infty} \frac{1}{log\, x - log\, \sigma} = 0.$$

Thus, Pareto(I)(2,$\sigma$) $\in$ DAN. In sum, Pareto(I)($\alpha$,$\sigma$) $\in$ DAN for $\alpha \ge 2$.

Let

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}, \;\; S_n(X) = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}}, \;\; \mu = E(X), \tag{3.7}$$

and

$$T_n(X) = \frac{\sum_{i=1}^{n} X_i}{S_n(X)\sqrt{n}}, \tag{3.8}$$

where the latter is the Student $t$-statistic.

It is known that

$$T_n(X-\mu) = \frac{\sum_{i=1}^{n}(X_i-\mu)}{S_n(X)\sqrt{n}} \tag{3.9}$$

converges in distribution to $N(0,1)$ when $0 < Var(X) < \infty$. Moreover, Giné et al. (1997) proved that

$$T_n(X-\mu) \xrightarrow[n\to\infty]{D} N(0,1) \;\Leftrightarrow\; X \in \text{ DAN} \quad \text{and} \quad E(X)=\mu. \tag{3.10}$$

In other words, (3.10) shows that the condition of $0 < Var(X) < \infty$ can be relaxed to having $X \in$ DAN, and the latter condition is both necessary and sufficient for $T_n(X-\mu)$ to be asymptotically $N(0,1)$.

We will now introduce our first asymptotic CI that is based on the asymptotic normality of $T_n(X-\mu)$ in (3.10). Let $X \in$ Pareto(I)($\alpha$,$\sigma$) with $\alpha \geq 2$, then $E(X) = \alpha\sigma/(\alpha-1)$ and (3.10) implies

$$\frac{\sum_{i=1}^{n}(X_i - \frac{\alpha\sigma}{\alpha-1})}{S_n(X)\sqrt{n}} \xrightarrow[n\to\infty]{D} N(0,1). \tag{3.11}$$

Having (3.11), we can easily build an asymptotic CI for $\alpha$. For $0 < \gamma < 1$, let $z_{\frac{\gamma}{2}}$ be such that $P(|N(0,1)| < z_{\frac{\gamma}{2}}) = 1-\gamma$. From (3.11),

$$P\Big(-z_{\frac{\gamma}{2}} \leq \frac{\sum_{i=1}^{n}(X_i - \frac{\alpha\sigma}{\alpha-1})}{S_n(X)\sqrt{n}} \leq z_{\frac{\gamma}{2}}\Big)$$

$$= P\Big(\frac{\sum_{i=1}^{n} X_i - z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} \le \frac{\alpha}{\alpha-1} \le \frac{\sum_{i=1}^{n} X_i + z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma}\Big)$$

$$= P\Big(\frac{\sum_{i=1}^{n} X_i - z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} - 1 \le \frac{1}{\alpha-1} \le \frac{\sum_{i=1}^{n} X_i + z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} - 1\Big)$$

$$\xrightarrow[n\to\infty]{} 1-\gamma, \tag{3.12}$$

and we have

$$P\Big(\Big(\frac{\sum_{i=1}^{n} X_i + z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} - 1\Big)^{-1} + 1 \le \alpha \le \Big(\frac{\sum_{i=1}^{n} X_i - z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} - 1\Big)^{-1} + 1\Big)$$

$$\xrightarrow[n\to\infty]{} 1-\gamma, \tag{3.13}$$

leading to the following $100(1-\gamma)\%$ asymptotic CI for $\alpha$ :

$$\left[\left(\frac{\sum_{i=1}^{n} X_i + z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} - 1\right)^{-1} + 1, \left(\frac{\sum_{i=1}^{n} X_i - z_{\frac{\gamma}{2}} S_n(X)\sqrt{n}}{n\sigma} - 1\right)^{-1} + 1\right], \tag{3.14}$$

which is valid for $\alpha \ge 2$.

Before we proceed, we have to note that to flip the last inequality in (3.12) to arrive to (3.13), we used the fact that its lower bound converges in probability to $E(X)/\sigma - 1 = 1/(\alpha-1) > 0$ (by, for example, Remark 2 and Example 1 of Martsynyuk (2013) and the weak law of large numbers (WLLN) that guarantee that $S_n(X)/\sqrt{n} \xrightarrow{P} 0$ and $\bar{X}_n \to E(X)$ as $n \to \infty$). Hence the probability of this bound being non-positive converges to 0, as $n \to \infty$, and can be ignored. The latter convergence can be seen from the following simple proposition that will help us to ignore similar asymptotically zero probabilities in the future as well.

**Proposition 3.1.** *Let* $\{X_n\}_{n\geq 1}$ *be a sequence of r.v.'s such that* $X_n \xrightarrow{P} c > 0$*, as* $n \to \infty$*. Then,* $P(X_n \leq 0) \to 0$*, as* $n \to \infty$*.*

*Proof.* Let $X_n \xrightarrow{P} c$, as $n \to \infty$. Then,

$$P\Big(|X_n - c| \leq \frac{c}{2}\Big) \to 1, \quad n \to \infty.$$

But,

$$P\Big(|X_n - c| \leq \frac{c}{2}\Big) \leq P\Big(\frac{c}{2} \leq X_n\Big) \leq P(0 < X_n).$$

Hence,

$$P(0 < X_n) \to 1, \quad \text{or} \quad P(X_n \leq 0) \to 0, \quad n \to \infty.$$

□

Let's look at an $Exp(\alpha)$ counterpart of the CI in (3.14) that uses

$$\left\{log\left(\frac{X_i}{\sigma}\right)\right\}_{1\leq i\leq n} =: \{Y_i\}_{1\leq i\leq n} \tag{3.15}$$

instead of the original sample $X_1, ..., X_n$ (cf. (2.3)). We note that since Var($Y_1$) $= \alpha^{-2} < \infty$ for all $\alpha > 0$, using the convergence in (3.10) for the exponential r.v.'s $Y_i$'s relieves us of the restriction $\alpha \geq 2$ required for the CI in (3.14). Thus, we have

$$\frac{\sum_{i=1}^{n}(Y_i - \frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \xrightarrow[n\to\infty]{D} N(0,1), \quad n \to \infty, \tag{3.16}$$

with $S_n(Y)$ as in (3.7). Since

$$\begin{aligned}
&P\left(-z_{\frac{\gamma}{2}} \leq \frac{\sum_{i=1}^{n}(Y_i - \frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \leq z_{\frac{\gamma}{2}}\right) \\
=&P\left(-z_{\frac{\gamma}{2}} S_n(Y)\sqrt{n} - \sum_{i=1}^{n} Y_i \leq -\sum_{i=1}^{n}\frac{1}{\alpha} \leq z_{\frac{\gamma}{2}} S_n(Y)\sqrt{n} - \sum_{i=1}^{n} Y_i\right) \\
=&P\left(\frac{\sum_{i=1}^{n} Y_i - z_{\frac{\gamma}{2}} S_n(Y)\sqrt{n}}{n} \leq \frac{1}{\alpha} \leq \frac{\sum_{i=1}^{n} Y_i + z_{\frac{\gamma}{2}} S_n(Y)\sqrt{n}}{n}\right),
\end{aligned} \tag{3.17}$$

we obtain the following $100(1-\gamma)\%$ asymptotic CI for $\alpha > 0$ :

$$AI_1 := \left[\frac{n}{\sum_{i=1}^{n} Y_i + z_{\frac{\gamma}{2}} S_n(Y)\sqrt{n}}, \frac{n}{\sum_{i=1}^{n} Y_i - z_{\frac{\gamma}{2}} S_n(Y)\sqrt{n}}\right], \tag{3.18}$$

where concluding $AI_1$ from (3.16) and (3.17) is similar to arriving from (3.12) to (3.14) via (3.13) and is based on applying Proposition 3.1 again.

In sum, both (3.14) and $AI_1$ of (3.18) are based on asymptotic normality of the Student $t$-statistic. However, as compared to the asymptotic CI in (3.14), $AI_1$ has no restriction on $\alpha$ and also performs better in terms of both the empirical expected length and finite-sample coverage probability, as illustrated in our Table 3.1, where $\hat{r}_{(3.14)} = \widehat{E(\text{Length of }(3.14))/E(\text{Length of } AI_1)}$, with $\widehat{E(\text{Length of } (3.14))}$ defined similarly to $\widehat{E(\text{Length of } AI_k)}$ of (3.2), and $\widehat{CP}_{(3.14)}$ defined like $\widehat{CP}_{AI_k}$ of (3.1). Therefore, in what follows, we will not consider (3.14) anymore and compare the rest of the CI's obtained in this chapter, one-by-one, to $AI_1$.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| | 50 | 1.330(0.755,0.879) | 1.619(0.798,0.929) | 1.838(0.839,0.955) |
| $\alpha$=2 | 100 | 1.165(0.787,0.892) | 1.409(0.830,0.936) | 1.732(0.877,0.972) |
| | 300 | 1.217(0.822,0.896) | 1.327(0.876,0.946) | 1.363(0.915,0.977) |
| | 1000 | 1.301(0.839,0.895) | 1.678(0.895,0.950) | 1.369(0.934,0.976) |
| | 50 | 1.039(0.824,0.875) | 1.729(0.881,0.930) | 1.445(0.910,0.958) |
| $\alpha$=3 | 100 | 1.050(0.853,0.891) | 1.068(0.901,0.940) | 1.178(0.941,0.973) |
| | 300 | 1.070(0.876,0.894) | 1.078(0.925,0.947) | 1.107(0.956,0.977) |
| | 1000 | 1.103(0.886,0.900) | 1.106(0.937,0.946) | 1.106(0.969,0.978) |
| | 50 | 1.013(0.856,0.877) | 1.037(0.906,0.926) | 1.097(0.940,0.961) |
| $\alpha$=5 | 100 | 1.014(0.877,0.891) | 1.022(0.928,0.941) | 1.035(0.958,0.970) |
| | 300 | 1.022(0.890,0.898) | 1.024(0.937,0.946) | 1.028(0.971,0.976) |
| | 1000 | 1.028(0.896,0.898) | 1.029(0.948,0.950) | 1.029(0.977,0.979) |

Table 3.1: (3.14) vs $AI_1$
$\hat{r}_{(3.14)}\,(\widehat{CP}_{(3.14)}, \widehat{CP}_{AI_1})$

## 3.2 Functional CLT based confidence intervals

The CI's for $\alpha$ in the previous section were based on the asymptotic normality of the Student $t$-statistic. In this section, we will consider convergence in distribution of r.v.'s that are based on the corresponding Student process. The Student process in $D[0,1]$, the space of real-valued functions on [0,1] that are right-continuous and have left-hand limits, is defined as follows:

$$T_n^t(X) = \frac{\sum_{i=1}^{[nt]} X_i}{S_n(X)\sqrt{n}}, \qquad \text{for } 0 \leq t \leq 1, \tag{3.19}$$

where $\sum_{i=1}^{0} X_i := 0$. We note that $T_n^1(X)$ is the Student $t$-statistic.

Let $\{W(t), 0 \leq t \leq 1\}$ be a standard Wiener process, $\rho$ be the sup-norm metric in $D[0,1]$ and $\mathscr{D}$ be the $\sigma$-field generated by the finite-dimensional subsets of $D[0,1]$.

Generalizing (3.10), Csörgő et al. (2003) proved:

$$X \in \text{DAN} \quad \text{and} \quad E(X) = \mu \ \Leftrightarrow \ h(T_n^t(X-\mu)) \xrightarrow[n\to\infty]{D} h(W(t)), \tag{3.20}$$

for all functionals $h : D[0,1] \to \mathbb{R}$ that are $\mathscr{D}$-measurable and $\rho$-continuous, or $\rho$-continuous except at points forming a set of Wiener measure zero on $(D[0,1], \mathscr{D})$. The convergence in distribution in (3.20) is called the functional CLT (FCLT).

As a consequence of (3.20), Martsynyuk and Tuzov (2016) considered convergence in distribution of five special functionals of the Student process and derived respective asymptotic CI's for $\mu = E(X)$, assuming that $X \in \text{DAN}$. They concluded that the obtained CI's may present reasonable alternatives to a classical asymptotic CI that follows simply from the asymptotic normality of the Student $t$-statistic in (3.10), due to them having higher finite-sample coverage probabilities, or shorter expected lengths.

In this section, inspired by Martsynyuk and Tuzov (2016), we adapt convergence in distribution of (3.20) of three of their functionals to constructing asymptotic CI's for $\alpha$ of Pareto(I)($\alpha$,$\sigma$), assuming that $\sigma$ is known. This method of building asymptotic CI's for $\alpha$ appears to be new. In fact, throughout the section, instead of dealing with the Pareto(I)($\alpha$,$\sigma$) sample $\{X_i\}_{1\leq i\leq n}$, we will apply (3.20) to its $Exp(\alpha)$ counterpart $\{Y_i\}_{1\leq i\leq n}$ of (3.15). This transformation has its advantages for our FCLT based CI's in the same way that $AI_1$ has over (3.14). First of all, while Pareto(I)$(\alpha,\sigma) \in$ DAN for $\alpha \geq 2$, $Exp(\alpha) \in$ DAN for all possible values of the tail index $\alpha > 0$. Secondly, in our preliminary studies (not included here), we constructed CI's based on the asymptotically pivotal quantities of this section both for the Pareto(I)$(\alpha,\sigma)$ and $Exp(\alpha)$ samples and, using simulations, concluded that all the $Exp(\alpha)$ based CI's perform better in terms of both the finite-sample coverage probability and expected length.

## CI based on $h_1$

Let $t_0 \in$ (0,1] be fixed and consider the following special functional for $f(t) \in D[0,1]$:

$$h_1(f(t)) = f(t_0). \tag{3.21}$$

By applying the FCLT of (3.20) with $h = h_1$, we have:

$$\frac{\sum_{i=1}^{[nt_0]}(Y_i - \frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \xrightarrow[n\to\infty]{D} N(0, t_0), \tag{3.22}$$

where $S_n(Y)$ is as in (3.7). Clearly, when $t_0$=1, (3.22) becomes (3.16) from section 3.1.

For $0< \gamma <1$,

$$P\Big( - z_{\frac{\gamma}{2}}\sqrt{t_0} \leq \frac{\sum_{i=1}^{[nt_0]}(Y_i - \frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \leq z_{\frac{\gamma}{2}}\sqrt{t_0}\Big)$$

$$=P\Big( - z_{\frac{\gamma}{2}}S_n(Y)\sqrt{nt_0} - \sum_{i=1}^{[nt_0]} Y_i \leq -\frac{[nt_0]}{\alpha} \leq z_{\frac{\gamma}{2}}S_n(Y)\sqrt{nt_0} - \sum_{i=1}^{[nt_0]} Y_i\Big)$$

$$=P\Big(\frac{\sum_{i=1}^{[nt_0]} Y_i - z_{\frac{\gamma}{2}}S_n(Y)\sqrt{nt_0}}{[nt_0]} \leq \frac{1}{\alpha} \leq \frac{\sum_{i=1}^{[nt_0]} Y_i + z_{\frac{\gamma}{2}}S_n(Y)\sqrt{nt_0}}{[nt_0]}\Big). \tag{3.23}$$

In view of the WLLN and Proposition 3.1, we can solve the last inequality for $\alpha$ on a set whose probability converges to one as $n \to \infty$, and conclude the following $(1-\gamma)100\%$ asymptotic CI for $\alpha$:

$$AI_2 := \left[\frac{[nt_0]}{\sum_{i=1}^{[nt_0]} Y_i + z_{\frac{\gamma}{2}}S_n(Y)\sqrt{nt_0}}, \frac{[nt_0]}{\sum_{i=1}^{[nt_0]} Y_i - z_{\frac{\gamma}{2}}S_n(Y)\sqrt{nt_0}}\right]. \tag{3.24}$$

Now, before comparing $AI_2$ to $AI_1$ of (3.18) numerically, we are left with the problem of choosing the value of $t_0$ in (3.24). Using preliminary simulations (not included here), we saw that higher values of $t_0$ led to $AI_2$ being shorter on average, while lower values of $t_0$ increased the finite-sample coverage probability of $AI_2$. As a compromise solution, we choose $t_0 = 0.9$.

From Table 3.2 on the performance of $AI_2$ versus that of $AI_1$, whose entries are in the form of $\hat{r}_2\,(\widehat{CP}_{AI_2}, \widehat{CP}_{AI_1})$, we can see that the ratio $\hat{r}_2$ of the empirical expected lengths is somewhat larger than 1, implying that $AI_2$ is longer than $AI_1$ on average. The ratio tends to be a bit larger for the larger values of the confidence level $1-\gamma$. However, $AI_2$ mostly has slightly higher finite-sample coverage probabilities than $AI_1$.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| $\alpha$=0.5 | 50 | 1.071(0.886,0.884) | 1.074(0.931,0.930) | 1.080(0.965,0.962) |
| | 100 | 1.062(0.887,0.887) | 1.063(0.938,0.935) | 1.065(0.973,0.974) |
| | 300 | 1.057(0.897,0.898) | 1.057(0.948,0.947) | 1.057(0.977,0.976) |
| | 1000 | 1.054(0.898,0.897) | 1.055(0.949,0.951) | 1.055(0.977,0.976) |
| $\alpha$=1.5 | 50 | 1.069(0.878,0.881) | 1.072(0.930,0.926) | 1.081(0.961,0.960) |
| | 100 | 1.062(0.893,0.894) | 1.063(0.944,0.942) | 1.065(0.972,0.968) |
| | 300 | 1.057(0.898,0.895) | 1.057(0.945,0.945) | 1.058(0.979,0.978) |
| | 1000 | 1.055(0.902,0.904) | 1.055(0.947,0.946) | 1.055(0.980,0.979) |
| $\alpha$=2 | 50 | 1.070(0.883,0.880) | 1.074(0.932,0.930) | 1.081(0.963,0.959) |
| | 100 | 1.062(0.893,0.893) | 1.062(0.937,0.934) | 1.065(0.972,0.972) |
| | 300 | 1.056(0.901,0.901) | 1.057(0.945,0.947) | 1.058(0.976,0.976) |
| | 1000 | 1.054(0.893,0.891) | 1.055(0.946,0.948) | 1.055(0.977,0.978) |
| $\alpha$=3 | 50 | 1.072(0.880,0.877) | 1.075(0.931,0.932) | 1.080(0.961,0.959) |
| | 100 | 1.062(0.892,0.892) | 1.063(0.944,0.940) | 1.066(0.969,0.968) |
| | 300 | 1.056(0.895,0.894) | 1.057(0.950,0.947) | 1.058(0.977,0.977) |
| | 1000 | 1.054(0.904,0.902) | 1.055(0.945,0.945) | 1.055(0.979,0.981) |
| $\alpha$=5 | 50 | 1.068(0.884,0.880) | 1.074(0.927,0.923) | 1.078(0.963,0.960) |
| | 100 | 1.062(0.892,0.892) | 1.064(0.941,0.940) | 1.065(0.969,0.970) |
| | 300 | 1.057(0.903,0.900) | 1.057(0.943,0.942) | 1.058(0.976,0.977) |
| | 1000 | 1.055(0.900,0.900) | 1.055(0.947,0.949) | 1.055(0.976,0.978) |

Table 3.2: $AI_2$ vs $AI_1$
$\hat{r}_2\,(\widehat{CP}_{AI_2}, \widehat{CP}_{AI_1})$

We note that the R code for producing Table 3.2 and the upcoming Tables 3.3 - 3.8 on $AI_k$ versus $AI_1$, $k=\overline{3,8}$, is written in such a way that it ignores all rarely occurring samples that violate the conditions for deriving $AI_1$ - $AI_8$, like for example the condition for $AI_2$ that the lower bound of the inequality in (3.23) is positive.

## CI based on $h_2$

Now, on applying the FCLT with the functional

$$h_2(f(t)) = \sup_{0\leq t\leq 1} |f(t)|, \tag{3.25}$$

where $f(t) \in D[0,1]$, we have:

$$\max_{1\leq k\leq n} \frac{|\sum_{i=1}^{k}(Y_i - \frac{1}{\alpha})|}{S_n(Y)\sqrt{n}} \xrightarrow[n\to\infty]{D} \sup_{0\leq t\leq 1} |W(t)|. \tag{3.26}$$

We note that for $0 < \gamma < 1$, the values of $b$ such that

$$P\left(\sup_{0\leq t\leq 1} |W(t)| \leq b\right) = 1 - \gamma \tag{3.27}$$

were tabulated in Csörgő and Horváth (1985) and can be used in construction of a CI based on (3.26).

From (3.26),

$$P\Big(\max_{1\leq k\leq n} \frac{|\sum_{i=1}^{k}(Y_i - \frac{1}{\alpha})|}{S_n(Y)\sqrt{n}} \leq b\Big) = P\Big(\bigcap_{k=1}^{n} \Big\{\frac{|\sum_{i=1}^{k}(Y_i - \frac{1}{\alpha})|}{S_n(Y)\sqrt{n}} \leq b\Big\}\Big)$$

$$=P\Big(\bigcap_{k=1}^{n}\Big\{-b \leq \frac{\sum_{i=1}^{k}(Y_i - \frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \leq b\Big\}\Big)$$

$$=P\Big(\bigcap_{k=1}^{n}\Big\{\frac{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}{k} \leq \frac{1}{\alpha} \leq \frac{\sum_{i=1}^{k} Y_i + bS_n(Y)\sqrt{n}}{k}\Big\}\Big) \xrightarrow[n\to\infty]{} 1-\gamma. \tag{3.28}$$

To derive a $100(1-\gamma)\%$ CI from (3.28), we have to solve for $\alpha$ under the last probability sign. Let's first analyze the inequality

$$\frac{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}{k} \leq \frac{1}{\alpha} \leq \frac{\sum_{i=1}^{k} Y_i + bS_n(Y)\sqrt{n}}{k}, \tag{3.29}$$

keeping in mind that $\alpha > 0$. Clearly, since $Y_i$'s are positive, we have $(\sum_{i=1}^{k} Y_i + bS_n(Y)\sqrt{n})/k > 0$ for all $k$. If

$$\frac{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}{k} > 0 \quad \text{for all } k,$$

then the mutual solution of (3.29) for all $k$ yields the following CI for $\alpha$:

$$\left[\max_{1\leq k\leq n} \frac{k}{\sum_{i=1}^{k} Y_i + bS_n(Y)\sqrt{n}}, \min_{1\leq k\leq n} \frac{k}{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}\right]. \tag{3.30}$$

However, there may exist some indices $k$ for which

$$\frac{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}{k} \leq 0.$$

For such indices, (3.29) is equivalent to

$$\alpha \in \left[ \frac{k}{\sum_{i=1}^{k} Y_i + bS_n(Y)\sqrt{n}}, \infty \right). \tag{3.31}$$

However,

$$P\Big( \bigcap_{k=1}^{n} \Big\{ \frac{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}{k} \le 0 \Big\} \Big) \le P\Big( \frac{\sum_{i=1}^{n} Y_i - bS_n(Y)\sqrt{n}}{n} \le 0 \Big) \xrightarrow[n\to\infty]{} 0, \tag{3.32}$$

due to Proposition 3.1 and the WLLN that implies that $(\sum_{i=1}^{n} Y_i - bS_n(Y)\sqrt{n})/n$ $\xrightarrow{P} E(Y_1) = 1/\alpha > 0$. In other words, with probability approaching 1 as $n \to \infty$, we do not have (3.31) for *all* $k$, $1 \le k \le n$. Therefore, (3.28) leads to a finite $100(1-\gamma)\%$ CI for $\alpha$:

$$AI_3 :=$$

$$\left[ \max_{1\le k\le n} \frac{k}{\sum_{i=1}^{k} Y_i + bS_n(Y)\sqrt{n}}, \min_{\left\{k:\, \frac{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}}{k} > 0,\, 1\le k\le n\right\}} \frac{k}{\sum_{i=1}^{k} Y_i - bS_n(Y)\sqrt{n}} \right]. \tag{3.33}$$

The comparative simulation performance of $AI_3$ versus $AI_1$ is included in Table 3.3. Table 3.3 shows that the ratio $\hat{r}_3$ of the empirical expected lengths is larger than 1, implying that $AI_3$ is longer than $AI_1$ on average. The ratio decreases as $n$ increases or as confidence level increases. In terms of the finite-sample coverage probability, $AI_3$ performs better than $AI_1$, and the difference $\widehat{CP}_{AI_3} - \widehat{CP}_{AI_1}$ ranges from -0.1% to 2%, being higher when $n = 50$ or $n = 100$ and when the confidence level $1-\gamma$ is smaller.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| | 50 | 1.116(0.903,0.884) | 1.100(0.938,0.930) | 1.089(0.967,0.962) |
| $\alpha$=0.5 | 100 | 1.089(0.900,0.887) | 1.073(0.943,0.935) | 1.063(0.978,0.974) |
| | 300 | 1.064(0.906,0.898) | 1.055(0.953,0.947) | 1.044(0.978,0.976) |
| | 1000 | 1.051(0.905,0.897) | 1.043(0.955,0.951) | 1.034(0.976,0.976) |
| | 50 | 1.118(0.899,0.881) | 1.100(0.934,0.926) | 1.090(0.965,0.960) |
| $\alpha$=1.5 | 100 | 1.088(0.904,0.894) | 1.074(0.947,0.942) | 1.063(0.972,0.968) |
| | 300 | 1.064(0.906,0.895) | 1.054(0.948,0.945) | 1.044(0.980,0.978) |
| | 1000 | 1.051(0.907,0.904) | 1.041(0.946,0.946) | 1.033(0.979,0.979) |
| | 50 | 1.116(0.900,0.880) | 1.099(0.939,0.930) | 1.089(0.965,0.959) |
| $\alpha$=2 | 100 | 1.089(0.907,0.893) | 1.075(0.943,0.934) | 1.063(0.974,0.972) |
| | 300 | 1.065(0.908,0.901) | 1.054(0.950,0.947) | 1.044(0.977,0.976) |
| | 1000 | 1.051(0.895,0.891) | 1.043(0.950,0.948) | 1.035(0.979,0.978) |
| | 50 | 1.115(0.896,0.877) | 1.100(0.940,0.932) | 1.088(0.965,0.959) |
| $\alpha$=3 | 100 | 1.087(0.899,0.892) | 1.074(0.947,0.940) | 1.063(0.972,0.968) |
| | 300 | 1.065(0.902,0.894) | 1.053(0.954,0.947) | 1.045(0.979,0.977) |
| | 1000 | 1.052(0.910,0.902) | 1.041(0.948,0.945) | 1.035(0.980,0.981) |
| | 50 | 1.117(0.898,0.880) | 1.101(0.934,0.923) | 1.089(0.966,0.960) |
| $\alpha$=5 | 100 | 1.088(0.907,0.892) | 1.074(0.947,0.940) | 1.063(0.972,0.970) |
| | 300 | 1.064(0.909,0.900) | 1.053(0.949,0.942) | 1.044(0.979,0.977) |
| | 1000 | 1.051(0.907,0.900) | 1.042(0.950,0.949) | 1.034(0.979,0.978) |

Table 3.3: $AI_3$ vs $AI_1$
$\hat{r}_3\,(\widehat{CP}_{AI_3}, \widehat{CP}_{AI_1})$

## CI based on $h_3$

When we read the FCLT in (3.20) with an integral functional $h_3$, where

$$h_3(T_n^t(Y-\mu)) = \int_0^1 (T_n^t(Y-\mu))^m dt, \quad \text{for } m = 1 \text{ and } 2, \tag{3.34}$$

we have

$$\frac{1}{n}\sum_{k=1}^{n-1}\Big(\frac{\sum_{i=1}^{k}(Y_i-\frac{1}{\alpha})}{S_n(Y)\sqrt{n}}\Big)^m \xrightarrow[n\to\infty]{D} \int_0^1 W^m(t)dt. \tag{3.35}$$

We note that obtaining $100(1-\gamma)\%$ CI's for $\alpha$ from (3.35) requires knowing the exact or approximate distribution of the limiting r.v. $\int_0^1 W^m(t)dt$. It is well known that

$$\int_0^1 W(t)dt \sim N\Big(0,\frac{1}{3}\Big)$$

(cf. Pinsky and Karlin (2011, p. 442), for example), while the values of $b$ satisfying

$$P\Big(\int_0^1 W^2(t)dt \le b\Big) = 1-\gamma \tag{3.36}$$

were tabulated in Csörgő and Horváth (1985).

### (3.35) with $m = 1$

For $m = 1$, (3.35) becomes

$$\frac{1}{n}\sum_{k=1}^{n-1}\frac{\sum_{i=1}^{k}(Y_i-\frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \xrightarrow[n\to\infty]{D} N\Big(0,\frac{1}{3}\Big). \tag{3.37}$$

Let $0<\gamma<1$ and $a$ and $b$ be such that

$$P(a<N(0,1)<b) = 1-\gamma, \tag{3.38}$$

where $\gamma_1 + \gamma_2 = \gamma$. Then, (3.37) implies that

$$P\left(\frac{a}{\sqrt{3}} \le \frac{1}{n}\sum_{k=1}^{n-1}\frac{\sum_{i=1}^{k}(Y_i - \frac{1}{\alpha})}{S_n(Y)\sqrt{n}} \le \frac{b}{\sqrt{3}}\right) =$$

$$P\left(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i - \frac{bS_n(Y)\sqrt{n}}{\sqrt{3}} \le \frac{1}{\alpha n}\frac{n(n-1)}{2} \le \frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i - \frac{aS_n(Y)\sqrt{n}}{\sqrt{3}}\right) =$$

$$P\left(\frac{2}{n-1}\Big(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i - \frac{bS_n(Y)\sqrt{n}}{\sqrt{3}}\Big) \le \frac{1}{\alpha} \le \frac{2}{n-1}\Big(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i - \frac{aS_n(Y)\sqrt{n}}{\sqrt{3}}\Big)\right)$$

$$\xrightarrow[n\to\infty]{} 1-\gamma, \tag{3.39}$$

leading to a $100(1-\gamma)\%$ asymptotic CI for $\alpha$:

$$AI_4 :=$$

$$\left[\frac{n-1}{2}\Big(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i - \frac{bS_n(Y)\sqrt{n}}{\sqrt{3}}\Big)^{-1}, \frac{n-1}{2}\Big(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i - \frac{aS_n(Y)\sqrt{n}}{\sqrt{3}}\Big)^{-1}\right]. \tag{3.40}$$

To check the validity of going from (3.39) to (3.40), we need to make sure that the probability of the lower (and hence the upper) bound of (3.39) being non-positive goes to 0, as $n \to \infty$. Since $(2bS_n(Y)\sqrt{n})/(\sqrt{3}(n-1)) \xrightarrow{P} 0$, it suffices to show that $(2\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i)/(n(n-1))$ is a consistent estimator of $1/\alpha$, which is positive. From (3.37), by using that $S_n(Y) \xrightarrow{P} 1/\alpha,\ n \to \infty$, and applying Slutsky's theorem,

$$\frac{1}{n}\sum_{k=1}^{n-1}\frac{\sum_{i=1}^{k}(Y_i-\frac{1}{\alpha})}{\sqrt{n}} \xrightarrow[n\to\infty]{D} N\left(0,\frac{1}{3\alpha^2}\right),$$

but

$$\frac{1}{n}\sum_{k=1}^{n-1}\frac{\sum_{i=1}^{k}(Y_i-\frac{1}{\alpha})}{\sqrt{n}} = \frac{\sqrt{n}}{2}\frac{n-1}{n}\Big(\frac{2\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i}{(n-1)n}-\frac{1}{\alpha}\Big)$$

and, applying Slutsky's theorem again, we get

$$\frac{2\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i}{(n-1)n}-\frac{1}{\alpha}\xrightarrow{P}0,\ \ n\to\infty, \tag{3.41}$$

that is $(2\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i)/(n(n-1))$ is indeed a consistent estimator of $1/\alpha$.

Now, let's look at the length of (3.40), which is

$$\frac{n-1}{2}\left[\left(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i-\frac{aS_n(Y)\sqrt{n}}{\sqrt{3}}\right)^{-1}-\left(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i-\frac{bS_n(Y)\sqrt{n}}{\sqrt{3}}\right)^{-1}\right] \tag{3.42}$$

$$=\frac{n-1}{2}\frac{(b-a)\frac{S_n(Y)\sqrt{n}}{\sqrt{3}}}{\left(\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i\right)^2-\frac{(b+a)S_n(Y)\sqrt{n}}{\sqrt{3}}\frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i+\frac{abS_n^2(Y)n}{3}}$$

$$=\frac{2}{n-1}\frac{(b-a)\frac{S_n(Y)\sqrt{n}}{\sqrt{3}}}{\left(\frac{2}{n(n-1)}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i\right)^2-\frac{2(b+a)S_n(Y)\sqrt{n}}{\sqrt{3}(n-1)}\left(\frac{2}{(n-1)n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}Y_i\right)+\frac{4abS_n^2(Y)n}{3(n-1)^2}}$$

$$=\frac{2\,(b-a)\,\sqrt{n}}{\sqrt{3}\,(n-1)}\frac{1/\alpha+op(1)}{(1/\alpha)^2+op(1)},$$

on account of (3.41) and the WLLN, where $op(1)$ stands for a sequence of r.v.'s that converges in probability to 0, as $n\to\infty$. Thus, for large enough

$n$, the length of $AI_4$ is essentially proportional to $b - a$, where $a$ and $b$ are as in (3.38). Therefore, it is reasonable to select $-a = b = z_{\gamma/2}$ in order to approximately minimize (3.42) (cf. (2.10)). Naturally, this choice of $a$ and $b$ does not guarantee higher finite-sample coverage probabilities for $AI_4$.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| $\alpha$=0.5 | 50 | 1.237(0.890,0.884) | 1.258(0.937,0.930) | 1.281(0.971,0.962) |
| | 100 | 1.194(0.891,0.887) | 1.199(0.942,0.935) | 1.206(0.977,0.974) |
| | 300 | 1.167(0.896,0.898) | 1.169(0.948,0.947) | 1.169(0.978,0.976) |
| | 1000 | 1.157(0.896,0.897) | 1.159(0.953,0.951) | 1.159(0.976,0.976) |
| $\alpha$=1.5 | 50 | 1.240(0.888,0.881) | 1.257(0.931,0.926) | 1.281(0.966,0.960) |
| | 100 | 1.194(0.892,0.894) | 1.197(0.945,0.942) | 1.206(0.973,0.968) |
| | 300 | 1.167(0.901,0.895) | 1.167(0.946,0.945) | 1.170(0.978,0.978) |
| | 1000 | 1.158(0.903,0.904) | 1.159(0.946,0.946) | 1.160(0.980,0.979) |
| $\alpha$=2 | 50 | 1.239(0.888,0.880) | 1.256(0.935,0.930) | 1.279(0.968,0.959) |
| | 100 | 1.191(0.899,0.893) | 1.195(0.940,0.934) | 1.206(0.974,0.972) |
| | 300 | 1.165(0.898,0.901) | 1.168(0.948,0.947) | 1.170(0.976,0.976) |
| | 1000 | 1.158(0.893,0.891) | 1.158(0.947,0.948) | 1.159(0.977,0.978) |
| $\alpha$=3 | 50 | 1.239(0.882,0.877) | 1.254(0.939,0.932) | 1.282(0.967,0.959) |
| | 100 | 1.192(0.891,0.892) | 1.197(0.943,0.940) | 1.206(0.973,0.968) |
| | 300 | 1.166(0.895,0.894) | 1.167(0.949,0.947) | 1.171(0.979,0.977) |
| | 1000 | 1.158(0.899,0.902) | 1.159(0.947,0.945) | 1.159(0.980,0.981) |
| $\alpha$=5 | 50 | 1.234(0.893,0.880) | 1.255(0.936,0.923) | 1.282(0.967,0.960) |
| | 100 | 1.194(0.893,0.892) | 1.198(0.944,0.940) | 1.207(0.972,0.970) |
| | 300 | 1.167(0.902,0.900) | 1.168(0.945,0.942) | 1.170(0.975,0.977) |
| | 1000 | 1.159(0.902,0.900) | 1.158(0.946,0.949) | 1.160(0.979,0.978) |

Table 3.4: $AI_4$ vs $AI_1$
$\hat{r}_4\,(\widehat{CP}_{AI_4}, \widehat{CP}_{AI_1})$

In Table 3.4, we can see that the ratio $\hat{r}_4$ of the empirical expected lengths is bigger than 1, implying that $AI_4$ is longer than $AI_1$ on average. The ratio decreases as $n$ increases, and increases as the confidence level increases. The empirical finite-sample coverage probability of $AI_4$ is mostly higher than that of $AI_1$. The difference $\widehat{CP}_{AI_4}$-$\widehat{CP}_{AI_1}$ ranges from -0.3% to 1.3%.

### (3.35) with $m = 2$

We now consider (3.35) with $m = 2$.

With $b$ as in (3.36),

$$P\left(\frac{1}{n}\sum_{k=1}^{n-1}\Big(\frac{\sum_{i=1}^{k}(Y_i-\frac{1}{\alpha})}{S_n(Y)\sqrt{n}}\Big)^2 \le b\right)$$

$$= P\left(\sum_{k=1}^{n-1}\Big(\sum_{i=1}^{k}Y_i\Big)^2 - \frac{2}{\alpha}\sum_{k=1}^{n-1}k\sum_{i=1}^{k}Y_i + \frac{1}{\alpha^2}\sum_{k=1}^{n-1}k^2 \le bS_n^2(Y)n^2\right)$$

$$= P\left(\frac{C_n}{\alpha^2} - \frac{2}{\alpha}\sum_{k=1}^{n-1}k\sum_{i=1}^{k}Y_i + \sum_{k=1}^{n-1}\Big(\sum_{i=1}^{k}Y_i\Big)^2 - bS_n^2(Y)n^2 \le 0\right), \qquad (3.43)$$

where

$$C_n := \sum_{k=1}^{n-1}k^2 = \frac{n(n-1)(2n-1)}{6}.$$

Whenever

$$D_n := \sqrt{\left(\sum_{k=1}^{n-1}k\sum_{i=1}^{k}Y_i\right)^2 - C_n\left(\sum_{k=1}^{n-1}\left(\sum_{i=1}^{k}Y_i\right)^2 - bS_n^2(Y)n^2\right)} > 0,$$

solving the quadratic inequality for $1/\alpha$ under the probability sign in (3.43) as

$$A_{1n} \le \frac{1}{\alpha} \le A_{2n},$$

with

$$A_{1n,2n} = \frac{\sum_{k=1}^{n-1} k \sum_{i=1}^{k} Y_i \mp \sqrt{D_n}}{C_n}, \tag{3.44}$$

leads to a $100(1-\gamma)\%$ asymptotic CI for $\alpha$:

$$AI_5 := \left[\frac{1}{A_{2n}}, \frac{1}{A_{1n}}\right]. \tag{3.45}$$

To justify the validity of $AI_5$, it needs to be shown that $P(D_n > 0) \to 1$ and $P(\sum_{k=1}^{n-1}(\sum_{i=1}^{k} Y_i)^2 - bS_n^2(Y)n^2 > 0) \to 1$ as $n \to \infty$ to ensure that we have two distinct positive zeroes for the quadratic function in $1/\alpha$ in (3.43) with probability approaching one, as $n \to \infty$. Indeed, the condition $D_n > 0$ guarantees the existence of two distinct zeroes, while the condition that the free term $\sum_{k=1}^{n-1}(\sum_{i=1}^{k} Y_i)^2 - bS_n^2(Y)n^2$ in (3.43) is positive guarantees that they are positive. We were unable to prove these convergences analytically, but preliminary simulations estimating $P(D_n > 0)$ and $P(\sum_{k=1}^{n-1}(\sum_{i=1}^{k} Y_i)^2 - bS_n^2(Y)n^2 > 0)$ for $n = 500, 1000, ...$ indicated convergence to 1 of these probabilities.

Table 3.5 on $AI_5$ versus $AI_1$, with the entries of the form $\hat{r}_5\,(\widehat{CP}_{AI_5}, \widehat{CP}_{AI_1})$, shows that $AI_5$ is longer than $AI_1$ on average. The ratio $\hat{r}_5$ decreases as $n$ increases, and increases as the confidence level increases. $AI_5$ has higher empirical finite-sample coverage probabilities than those of $AI_1$ in most cases, and the difference $\widehat{CP}_{AI_5}$-$\widehat{CP}_{AI_1}$ ranges from -0.4% to 1%.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| $\alpha$=0.5 | 50 | 1.155(0.890,0.884) | 1.166(0.935,0.930) | 1.178(0.968,0.962) |
| | 100 | 1.127(0.891,0.887) | 1.130(0.940,0.935) | 1.133(0.975,0.974) |
| | 300 | 1.109(0.897,0.898) | 1.112(0.948,0.947) | 1.111(0.978,0.976) |
| | 1000 | 1.103(0.898,0.897) | 1.104(0.952,0.951) | 1.105(0.977,0.976) |
| $\alpha$=1.5 | 50 | 1.156(0.885,0.881) | 1.165(0.931,0.926) | 1.178(0.965,0.960) |
| | 100 | 1.127(0.894,0.894) | 1.129(0.945,0.942) | 1.134(0.973,0.968) |
| | 300 | 1.109(0.902,0.895) | 1.110(0.945,0.945) | 1.111(0.979,0.978) |
| | 1000 | 1.104(0.900,0.904) | 1.104(0.944,0.946) | 1.105(0.979,0.979) |
| $\alpha$=2 | 50 | 1.155(0.885,0.880) | 1.165(0.933,0.930) | 1.178(0.966,0.959) |
| | 100 | 1.125(0.897,0.893) | 1.127(0.938,0.934) | 1.133(0.973,0.972) |
| | 300 | 1.108(0.900,0.901) | 1.110(0.946,0.947) | 1.111(0.976,0.976) |
| | 1000 | 1.103(0.891,0.891) | 1.104(0.947,0.948) | 1.105(0.976,0.978) |
| $\alpha$=3 | 50 | 1.155(0.881,0.877) | 1.164(0.935,0.932) | 1.179(0.965,0.959) |
| | 100 | 1.125(0.889,0.892) | 1.128(0.941,0.940) | 1.135(0.972,0.968) |
| | 300 | 1.109(0.895,0.894) | 1.109(0.950,0.947) | 1.112(0.977,0.977) |
| | 1000 | 1.104(0.902,0.902) | 1.104(0.946,0.945) | 1.105(0.979,0.981) |
| $\alpha$=5 | 50 | 1.151(0.888,0.880) | 1.165(0.933,0.923) | 1.178(0.967,0.960) |
| | 100 | 1.126(0.894,0.892) | 1.130(0.943,0.940) | 1.135(0.972,0.970) |
| | 300 | 1.109(0.902,0.900) | 1.110(0.944,0.942) | 1.111(0.975,0.977) |
| | 1000 | 1.104(0.903,0.900) | 1.104(0.946,0.949) | 1.105(0.980,0.978) |

Table 3.5: $AI_5$ vs $AI_1$
$\hat{r}_5\,(\widehat{CP}_{AI_5}, \widehat{CP}_{AI_1})$

## 3.3 Confidence intervals based on asymptotic normality of estimators for $\alpha$

In this section, we derive CI's for $\alpha$ from the asymptotic normality of the MLE for $\alpha$, the method of moments estimator (MME) for $\alpha$, and the generalized median estimator (GME) for $\alpha$ respectively in subsections 3.3.1, 3.3.2, and 3.3.3. All three CI's will be compared to $AI_1$ of (3.18).

### 3.3.1 Maximum likelihood estimator for $\alpha$

The MLE $\hat{\alpha}_{mle}$ for $\alpha$ as in (2.2) is well-known to be asymptotically normal (cf. Arnold (1983, p. 226), for example):

$$\frac{\sqrt{n}}{\alpha}(\hat{\alpha}_{mle} - \alpha) \xrightarrow[n\to\infty]{D} N(0,1). \tag{3.46}$$

Since

$$\begin{aligned} &P\left(-z_{\frac{\gamma}{2}} \le \frac{\sqrt{n}}{\alpha}(\hat{\alpha}_{mle} - \alpha) \le z_{\frac{\gamma}{2}}\right) \\ =&P\left(\frac{1}{\hat{\alpha}_{mle}}\left(1 - \frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}\right) \le \frac{1}{\alpha} \le \frac{1}{\hat{\alpha}_{mle}}\left(1 + \frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}\right)\right), \end{aligned} \tag{3.47}$$

where $P((1 - z_{\frac{\gamma}{2}}/\sqrt{n})/\hat{\alpha}_{mle} > 0) \to 1$, as $n \to \infty$, (3.46) leads to the following $100(1-\gamma)\%$ asymptotic CI for $\alpha$:

$$AI_6 := \left[\hat{\alpha}_{mle}\left(1 + \frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}\right)^{-1}, \hat{\alpha}_{mle}\left(1 - \frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}\right)^{-1}\right]. \tag{3.48}$$

We note in passing on why $-z_{\gamma/2}$ and $z_{\gamma/2}$ were chosen in (3.47). In general, $AI_6$ could look like $[\hat{\alpha}_{mle}(1+b/\sqrt{n})^{-1}, \hat{\alpha}_{mle}(1+a/\sqrt{n})^{-1}]$, with $a$ and $b$ satisfying (3.38). The length of such a CI would be

$$\hat{\alpha}_{mle}\frac{(b-a)}{\sqrt{n}}\frac{1}{1 + \frac{a+b}{\sqrt{n}} + \frac{ab}{n}}.$$

For large $n$, the whole expression is essentially proportional to $(b-a)$ and approximately minimized when $-a = b = z_{\gamma/2}$.

Fortunately, it is not difficult to obtain the closed form expression for the expected length of $AI_6$. Indeed, by using (2.8),

$$E(\text{Length of } AI_6) = E(\hat{\alpha}_{mle}) \left[ \left(1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right)^{-1} - \left(1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right)^{-1} \right]$$

$$= \frac{\alpha n}{n-1} \frac{2\, z_{\frac{\gamma}{2}} \sqrt{n}}{n - z_{\frac{\gamma}{2}}^2}. \tag{3.49}$$

We can think of a variation of (3.46). Since $\hat{\alpha}_{mle} \xrightarrow{P} \alpha$, $n \to \infty$, using Slutsky's theorem, we can replace $\alpha$ in the denominator in (3.46) with $\hat{\alpha}_{mle}$:

$$\frac{\sqrt{n}}{\hat{\alpha}_{mle}}(\hat{\alpha}_{mle} - \alpha) \xrightarrow[n\to\infty]{D} N(0,1). \tag{3.50}$$

Using (3.50), a $100(1-\gamma)\%$ asymptotic CI, denoted by $AI_7$, is easily derived:

$$AI_7 := \left[ \hat{\alpha}_{mle} - \frac{z_{\frac{\gamma}{2}} \hat{\alpha}_{mle}}{\sqrt{n}}, \hat{\alpha}_{mle} + \frac{z_{\frac{\gamma}{2}} \hat{\alpha}_{mle}}{\sqrt{n}} \right]. \tag{3.51}$$

The endpoints $\mp z_{\gamma/2}$ in (3.51) were chosen to minimize the length of $AI_7$, in view of (2.10) applied to the limiting $N(0,1)$ distribution. From (2.8), the expected length of $AI_7$ is:

$$E(\text{Length of } AI_7) = E\left[2\hat{\alpha}_{mle} \frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}\right] = \frac{2\alpha n}{n-1} \frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}.$$

We can also campare the expected lengths of $AI_6$ and $AI_7$:

$$\frac{E(\text{Length of } AI_6)}{E(\text{Length of } AI_7)} = \frac{\dfrac{2\alpha n}{n-1}\dfrac{z_{\frac{\gamma}{2}}\sqrt{n}}{n-z^2_{\frac{\gamma}{2}}}}{\dfrac{2\alpha n}{n-1}\dfrac{z_{\frac{\gamma}{2}}}{\sqrt{n}}} = \frac{n}{n-z^2_{\frac{\gamma}{2}}}.$$

For our settings with $\gamma \leq 0.1$ and $n \geq 50$, $AI_6$ is always a bit longer than $AI_7$.

We produce Tables 3.6 and 3.7 to compare the performance of $AI_6$ and $AI_7$ to that of $AI_1$.

From Table 3.6, $AI_6$ is a bit longer than $AI_1$ on average, less so as $n$ increases. The empirical finite-sample coverage probability of $AI_6$ is almost always higher than that of $AI_1$, with the difference $\widehat{CP}_{AI_6}$-$\widehat{CP}_{AI_1}$ ranging from -0.1% to 1.4% and being higher for smaller $n$'s, $n = 50$ and $n = 100$. Similarly to the expected lengths, the coverage probabilities of $AI_6$ and $AI_1$ become almost the same as $n$ increases.

From Table 3.7, we can see that the ratio $\hat{r}_7$ of the expected lengths of $AI_7$ and $AI_1$ is less than 1, implying that $AI_7$ is shorter than $AI_1$ on average. Although $\hat{r}_7$ increases as $n$ increases, a simulation run for $n$ as large as 30,000 (not included here) showed that $\hat{r}_7$ is still slightly less than 1 for such large $n$. The ratio $\hat{r}_7$ decreases as the confidence level increases. Moreover, the difference $\widehat{CP}_{AI_7}$-$\widehat{CP}_{AI_1}$ ranges from 0% to 3%. Thus, $AI_7$ performs better than $AI_1$ both in terms of the expected length and finite-sample coverage probability.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| $\alpha$=0.5 | 50 | 1.016(0.896,0.884) | 1.018(0.940,0.930) | 1.013(0.968,0.962) |
| | 100 | 1.010(0.897,0.887) | 1.011(0.942,0.935) | 1.008(0.978,0.974) |
| | 300 | 1.003(0.899,0.898) | 1.003(0.950,0.947) | 1.004(0.977,0.976) |
| | 1000 | 1.001(0.897,0.897) | 1.001(0.951,0.951) | 1.001(0.977,0.976) |
| $\alpha$=1.5 | 50 | 1.016(0.894,0.881) | 1.019(0.938,0.926) | 1.011(0.966,0.960) |
| | 100 | 1.010(0.900,0.894) | 1.006(0.946,0.942) | 1.009(0.973,0.968) |
| | 300 | 1.004(0.896,0.895) | 1.004(0.948,0.945) | 1.003(0.980,0.978) |
| | 1000 | 1.001(0.903,0.904) | 1.001(0.947,0.946) | 1.001(0.980,0.979) |
| $\alpha$=2 | 50 | 1.018(0.891,0.880) | 1.017(0.940,0.930) | 1.015(0.967,0.959) |
| | 100 | 1.010(0.901,0.893) | 1.010(0.941,0.934) | 1.008(0.975,0.972) |
| | 300 | 1.004(0.907,0.901) | 1.004(0.950,0.947) | 1.004(0.978,0.976) |
| | 1000 | 1.001(0.892,0.891) | 1.001(0.949,0.948) | 1.001(0.979,0.978) |
| $\alpha$=3 | 50 | 1.018(0.891,0.877) | 1.019(0.943,0.932) | 1.020(0.966,0.959) |
| | 100 | 1.011(0.900,0.892) | 1.010(0.947,0.940) | 1.008(0.972,0.968) |
| | 300 | 1.004(0.899,0.894) | 1.004(0.951,0.947) | 1.003(0.978,0.977) |
| | 1000 | 1.001(0.904,0.902) | 1.001(0.944,0.945) | 1.000(0.980,0.981) |
| $\alpha$=5 | 50 | 1.015(0.892,0.880) | 1.019(0.936,0.923) | 1.016(0.968,0.960) |
| | 100 | 1.009(0.903,0.892) | 1.008(0.947,0.940) | 1.011(0.974,0.970) |
| | 300 | 1.004(0.904,0.900) | 1.004(0.945,0.942) | 1.003(0.978,0.977) |
| | 1000 | 1.001(0.904,0.900) | 1.001(0.948,0.949) | 1.001(0.980,0.978) |

Table 3.6: $AI_6$ vs $AI_1$
$\hat{r}_6\,(\widehat{CP}_{AI_6}, \widehat{CP}_{AI_1})$

### 3.3.2 Method of moments estimator for $\alpha$

We can construct a CI for $\alpha$ from the asymptotic normality of the method of moments estimator (MME) $\hat{\alpha}_{mme}$ for $\alpha$, where

$$\hat{\alpha}_{mme} = \frac{\bar{X}_n}{\bar{X}_n - \sigma}, \tag{3.52}$$

obtained by equating the sample mean $\bar{X}_n$ to the expected value $\alpha\sigma/(\alpha-1)$ of Pareto(I)$(\alpha, \sigma)$ and then solving for $\alpha$.

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| | 50 | 0.961(0.904,0.884) | 0.940(0.950,0.930) | 0.904(0.980,0.962) |
| $\alpha$=0.5 | 100 | 0.982(0.904,0.887) | 0.972(0.948,0.935) | 0.954(0.982,0.974) |
| | 300 | 0.994(0.901,0.898) | 0.990(0.950,0.947) | 0.986(0.980,0.976) |
| | 1000 | 0.998(0.899,0.897) | 0.997(0.951,0.951) | 0.995(0.977,0.976) |
| | 50 | 0.961(0.900,0.881) | 0.941(0.949,0.926) | 0.902(0.980,0.960) |
| $\alpha$=1.5 | 100 | 0.982(0.903,0.894) | 0.967(0.954,0.942) | 0.955(0.980,0.968) |
| | 300 | 0.994(0.903,0.895) | 0.991(0.952,0.945) | 0.985(0.981,0.978) |
| | 1000 | 0.998(0.904,0.904) | 0.997(0.947,0.946) | 0.995(0.979,0.979) |
| | 50 | 0.963(0.902,0.880) | 0.939(0.949,0.930) | 0.905(0.981,0.959) |
| $\alpha$=2 | 100 | 0.982(0.901,0.893) | 0.971(0.950,0.934) | 0.953(0.980,0.972) |
| | 300 | 0.995(0.903,0.901) | 0.991(0.949,0.947) | 0.986(0.979,0.976) |
| | 1000 | 0.999(0.891,0.891) | 0.997(0.950,0.948) | 0.996(0.978,0.978) |
| | 50 | 0.963(0.907,0.877) | 0.941(0.952,0.932) | 0.909(0.979,0.959) |
| $\alpha$=3 | 100 | 0.984(0.903,0.892) | 0.971(0.954,0.940) | 0.953(0.978,0.968) |
| | 300 | 0.995(0.901,0.894) | 0.991(0.951,0.947) | 0.985(0.980,0.977) |
| | 1000 | 0.999(0.906,0.902) | 0.997(0.946,0.945) | 0.995(0.981,0.981) |
| | 50 | 0.960(0.904,0.880) | 0.940(0.949,0.923) | 0.906(0.977,0.960) |
| $\alpha$=5 | 100 | 0.981(0.905,0.892) | 0.969(0.952,0.940) | 0.956(0.979,0.970) |
| | 300 | 0.995(0.903,0.900) | 0.991(0.949,0.942) | 0.985(0.979,0.977) |
| | 1000 | 0.999(0.902,0.900) | 0.997(0.950,0.949) | 0.996(0.980,0.978) |

Table 3.7: $AI_7$ vs $AI_1$
$\hat{r}_7\,(\widehat{CP}_{AI_7}, \widehat{CP}_{AI_1})$

Using the WLLN, we can easily check that $\hat{\alpha}_{mme}$ is a weakly consistent estimator of $\alpha$:

$$\hat{\alpha}_{mme} = \frac{\bar{X}_n}{\bar{X}_n - \sigma} \xrightarrow{P} \frac{E[X]}{E[X]-\sigma} = \frac{\frac{\alpha\sigma}{\alpha-1}}{\frac{\alpha\sigma}{\alpha-1} - \sigma} = \frac{\frac{\alpha\sigma}{\alpha-1}}{\frac{\sigma}{\alpha-1}} = \alpha, \quad n \to \infty.$$

It is also known that $\hat{\alpha}_{mme}$ is asymptotically normal:

$$\sqrt{n}(\hat{\alpha}_{mme} - \alpha) \xrightarrow[n\to\infty]{D} N\left(0, \frac{\alpha(\alpha-1)^2}{\alpha-2}\right), \text{ for } \alpha > 2. \tag{3.53}$$

To verify (3.53), we first note that by the CLT,

$$\sqrt{n}\left(\bar{X}_n - \frac{\alpha\sigma}{\alpha-1}\right) \xrightarrow{D} N\left(0, \frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}\right), \quad n \to \infty. \tag{3.54}$$

This holds true when $Var(X)$ is finite, that is when $\alpha > 2$. Let's consider now function

$$g(x) = \frac{x}{x-\sigma}. \tag{3.55}$$

Since

$$\hat{\alpha}_{mme} = g(\bar{X}_n) \quad \text{and} \quad g\left(\frac{\alpha\sigma}{\alpha-1}\right) = \alpha,$$

and

$$g'(x) = -\frac{\sigma}{(x-\sigma)^2}, \quad g'\left(\frac{\alpha\sigma}{\alpha-1}\right) = -\frac{\sigma}{(\frac{\alpha\sigma}{\alpha-1}-\sigma)^2} = -\frac{(\alpha-1)^2}{\sigma} \neq 0,$$

and

$$Var(X)\left[g'\left(\frac{\alpha\sigma}{\alpha-1}\right)\right]^2 = \frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)} \cdot \frac{(\alpha-1)^4}{\sigma^2} = \frac{\alpha(\alpha-1)^2}{\alpha-2},$$

convergence in (3.53) is obtained from that in (3.54) by the delta method.

To use (3.53) for constructing an asymptotic CI for $\alpha$, we first estimate the asymptotic variance in (3.53) with $\hat{\alpha}_{mme}(\hat{\alpha}_{mme}-1)^2/(\hat{\alpha}_{mme}-2)$ (as leaving it unestimated makes it too difficult to solve for $\alpha$) and then obtain the

$100(1-\gamma)\%$ asymptotic CI $AI_8$ for $\alpha > 2$ using Slutsky's theorem:

$$AI_8 := \left[\hat{\alpha}_{mme} - \frac{z_{\frac{\gamma}{2}}\sqrt{\hat{\alpha}_{mme}}(\hat{\alpha}_{mme}-1)}{\sqrt{n(\hat{\alpha}_{mme}-2)}}, \hat{\alpha}_{mme} + \frac{z_{\frac{\gamma}{2}}\sqrt{\hat{\alpha}_{mme}}(\hat{\alpha}_{mme}-1)}{\sqrt{n(\hat{\alpha}_{mme}-2)}}\right]. \tag{3.56}$$

Comparing the finite-sample coverage probabilties and expected lengths of $AI_8$ and $AI_1$ in Table 3.8, we observe that the ratio $\hat{r}_8$ of the empirical expected lengths is bigger than 1, with a few exceptions for $\alpha = 5$ and $n = 50$ and 100, which implies that $AI_8$ is longer than $AI_1$ on average. It should be noted that the ratio increases as $n$ increases. We run an additional simulation (not included here) with $n$ as large as 30,000 and the result showed $\hat{r}_8$ stays around 1.155 and 1.033 respectively for $\alpha = 3$ and 5 for such large $n$. The ratio decreases as the confidence level increases. $AI_8$ has mostly higher empirical finite-sample coverage probabilities than those of $AI_1$, with the difference $\widehat{CP}_{AI_8}$-$\widehat{CP}_{AI_1}$ ranging from -0.1% to 8.3% and being higher for smaller $n$'s, but the difference decreases as $n$ increases.

### 3.3.3 Generalized median estimator for $\alpha$

Brazauskas and Serfling (2000a) studied asymptotic performance of several estimators of the tail index $\alpha$ of the Pareto(I)$(\alpha, \sigma)$ distribution. They used two measures of asymptotic performance of an estimator: its asymptotic relative efficiencey (ARE with respect to the MLE of $\alpha$) and robustness against

| Pareto(I)($\alpha$,1) | $n$ | $1-\gamma$ 0.9 | 0.95 | 0.98 |
|---|---|---|---|---|
| $\alpha$=3 | 50 | 1.137(0.960,0.877) | 1.114(0.984,0.932) | 1.077(0.994,0.959) |
| | 100 | 1.151(0.940,0.892) | 1.138(0.978,0.940) | 1.115(0.993,0.968) |
| | 300 | 1.153(0.918,0.894) | 1.150(0.964,0.947) | 1.142(0.984,0.977) |
| | 1000 | 1.155(0.912,0.902) | 1.153(0.952,0.945) | 1.150(0.984,0.981) |
| $\alpha$=5 | 50 | 0.997(0.912,0.880) | 0.977(0.953,0.923) | 0.941(0.982,0.960) |
| | 100 | 1.016(0.908,0.892) | 1.004(0.956,0.940) | 0.990(0.983,0.970) |
| | 300 | 1.028(0.906,0.900) | 1.025(0.948,0.942) | 1.018(0.980,0.977) |
| | 1000 | 1.032(0.899,0.900) | 1.030(0.950,0.949) | 1.029(0.979,0.978) |

Table 3.8: $AI_8$ vs $AI_1$
$\hat{r}_8\,(\widehat{CP}_{AI_8}, \widehat{CP}_{AI_1})$

upper outliers. ARE of an estimator is defined as the limiting ratio of the respective sample sizes at which this estimator and the MLE work equivalently in terms of the asymptotic variance criterion. In Brazauskas and Serfling (2000a), robustness of an estimator is defined via its upper breakdown point, characterized as the largest proportion of upper sample observations from Pareto(I)($\alpha$,$\sigma$) which may be taken to $\infty$ without taking the estimator to a limit not depending on the parameter being estimated. The MLE for $\alpha$ being efficient but nonrobust prompted the authors to look for alternative estimators for $\alpha$ with a relatively high efficiency and adequate robustness. In this regard, they introduced a new estimator for $\alpha$, called generalized median estimator (GME), and compared it with the MLE and other known estimators for $\alpha$, such as the ones obtained by the method of moments, trimming, least squares, and quantiles/percentile matching. Brazauskas and Serfling (2000a) concluded that the GME performs best in terms of both the robustness and efficiency. It led us to investigate the performance of the CI's for $\alpha$ that are based on the asymptotic normality of the GME.

First, let's consider how the GME for $\alpha$ was introduced in Brazauskas and Serfling (2000a).

Let $k \geq 1$ be an integer. Consider a kernel:

$$h_0(X_1, ..., X_k) = \Big(k^{-1}\sum_{j=1}^{k} \log X_j - \log\sigma\Big)^{-1}. \tag{3.57}$$

We observe that $h_0(X_1, ..., X_k)$ is the MLE for $\alpha$ based on the sample $X_1, ..., X_k$ of size $k$ (cf. (2.2)). Let $M_{2k}$ be the median of the $\chi^2_{2k}$ distribution. Then, using the fact that

$$2k\alpha\, h_0^{-1}(X_1, ..., X_k) \sim \chi^2_{2k}, \tag{3.58}$$

which corresponds to (2.1) with the sample size $k$, one can see that the r.v.

$$h(X_1, ..., X_k) = \frac{M_{2k}}{2k} h_0(X_1, ..., X_k) \tag{3.59}$$

has median $\alpha$. Then, the GME is defined by taking $\binom{n}{k}$ subsets of the sample $X_1, ..., X_n$, evaluating $h(\cdot)$ based on all those subsamples, and then taking the median of the obtained $h(\cdot)$'s:

$$\hat{\alpha}_{gme} = Median\Big\{h(X_{i_1}, ..., X_{i_k}), 1 \leq i_1 < ... < i_k \leq n\Big\}. \tag{3.60}$$

The values of $M_{2k}$ and $\frac{M_{2k}}{2k}$ for $k$=1:10 were tabulated in Brazauskas and Serfling (2000a). We note that when $k = n$, $\hat{\alpha}_{gme}$ is simply $(M_{2n}\,\hat{\alpha}_{mle})/2n$, with $\hat{\alpha}_{mle}$ of (2.2).

Brazauskas and Serfling (2000a) argued that $\hat{\alpha}_{gme}$ is asymptotically normal:

$$\frac{\sqrt{n}}{\alpha\sqrt{\gamma_k}}(\hat{\alpha}_{gme} - \alpha) \xrightarrow[n\to\infty]{D} N(0,1), \tag{3.61}$$

where $\gamma_k$ is a constant depending on $k$. They also tabulated the values of $\gamma_k$ for $k$=1:10.

From (3.61), we construct a $100(1-\gamma)\%$ CI for $\alpha$, $AI_9$, in the same manner as the previous CI's and using quantiles $-z_{\frac{\gamma}{2}}$ and $z_{\frac{\gamma}{2}}$ of $N(0,1)$. Since

$$P\left(-z_{\frac{\gamma}{2}} \le \frac{\sqrt{n}}{\alpha\sqrt{\gamma_k}}(\hat{\alpha}_{gme} - \alpha) \le z_{\frac{\gamma}{2}}\right)$$
$$=P\left(1 - \frac{z_{\frac{\gamma}{2}}\sqrt{\gamma_k}}{\sqrt{n}} \le \frac{\hat{\alpha}_{gme}}{\alpha} \le 1 + \frac{z_{\frac{\gamma}{2}}\sqrt{\gamma_k}}{\sqrt{n}}\right)$$

and, due to the consistency of $\hat{\alpha}_{gme}$, $P(\hat{\alpha}_{gme}(1 - z_{\gamma/2}\sqrt{\gamma_k}/\sqrt{n})^{-1} > 0) \to 1$, $n \to \infty$, we obtain

$$AI_9 := \left[\hat{\alpha}_{gme}\left(1 + \frac{z_{\frac{\gamma}{2}}\sqrt{\gamma_k}}{\sqrt{n}}\right)^{-1}, \hat{\alpha}_{gme}\left(1 - \frac{z_{\frac{\gamma}{2}}\sqrt{\gamma_k}}{\sqrt{n}}\right)^{-1}\right]. \tag{3.62}$$

We note that $\alpha$ in the denominator on the left-hand side of (3.61) can also be estimated with with $\hat{\alpha}_{gme}$, for example. Using such convergence, one can build one more GME based $100(1-\gamma)\%$ CI for $\alpha$:

$$AI_{10} := \left[\hat{\alpha}_{gme} - \frac{z_{\frac{\gamma}{2}}\hat{\alpha}_{gme}\sqrt{\gamma_k}}{\sqrt{n}}, \hat{\alpha}_{gme} + \frac{z_{\frac{\gamma}{2}}\hat{\alpha}_{gme}\sqrt{\gamma_k}}{\sqrt{n}}\right], \tag{3.63}$$

which is a slightly shorter CI on average than $AI_9$, since

$$\frac{E(\text{Length of } AI_9)}{E(\text{Length of } AI_{10})} = \frac{E(\hat{\alpha}_{gme})\dfrac{2z_{\gamma/2}\sqrt{n}\sqrt{\gamma_k}}{n - z^2_{\gamma/2}\gamma_k}}{E(\hat{\alpha}_{gme})\dfrac{2z_{\gamma/2}\sqrt{\gamma_k}}{\sqrt{n}}} = \frac{n}{n - z^2_{\gamma/2}\gamma_k} > 1.$$

However, the ratio of the expected lengths goes to 1, as $n \to \infty$.

As to comparing $AI_9$ and $AI_{10}$ to $AI_1$, we have to note that there was a computational cost problem related to evaluating $AI_9$ and $AI_{10}$ even for one repetition of the sample $X_1$,...,$X_n$, which really comes from the calculation of $\hat{\alpha}_{gme}$ based on evaluating $h(\cdot)$ of (3.59) for all $\binom{n}{k}$ subsamples $X_{i_1}$,...,$X_{i_k}$ of $X_1$,...,$X_n$. As a result, in our simulations, we had to compromise the sample size $n$, the number of repetitions of $X_1, ..., X_n$ and $k$ ($1 \leq k \leq 10$) while maintaining competitive expected lengths of $AI_9$ and $AI_{10}$. Accordingly, we decided on $k$=4, since the ARE of $\hat{\alpha}_{gme}$, decreasing with $k$, is still fairly high and equals 0.92 in this case (cf. Brazauskas and Serfling (2000a)). We chose $n = 50$, 100, 200, and 300, and the number of repetitions to be 500. Brazauskas and Serfling (2000a) discussed a possible remedy to this computational problem. They proposed to estimate $\hat{\alpha}_{gme}$ by taking only a big enough number $N$ of all $\binom{n}{k}$ required subsamples of $X_1, ..., X_n$. This way of computing is much more time efficient while still maintaining desired numerical accuracy. However, we did not apply it for our simulations to avoid the unspecified numerical inaccuracy.

Comparing $AI_9$ to $AI_1$ in Table 3.9, we can see that $AI_9$ is somewhat longer than $AI_1$ on average. The ratio $\hat{r}_9$ decreases as $n$ increases. $AI_9$ mostly has

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| $\alpha$=0.5 | 50 | 1.065(0.904,0.868) | 1.063(0.946,0.930) | 1.054(0.974,0.962) |
| | 100 | 1.043(0.904,0.906) | 1.052(0.946,0.934) | 1.051(0.972,0.976) |
| | 200 | 1.048(0.904,0.900) | 1.041(0.962, 0.960) | 1.043(0.982,0.982) |
| | 300 | 1.046(0.902,0.906) | 1.043(0.940,0.934) | 1.048(0.974,0.970) |
| $\alpha$=1.5 | 50 | 1.060(0.910,0.888) | 1.061(0.948,0.928) | 1.058(0.980,0.966) |
| | 100 | 1.052(0.878,0.874) | 1.054(0.938,0.926) | 1.056(0.978,0.964) |
| | 200 | 1.047(0.908,0.902) | 1.050(0.960, 0.946) | 1.050(0.970,0.976) |
| | 300 | 1.047(0.868,0.870) | 1.044(0.950,0.952) | 1.048(0.990,0.982) |
| $\alpha$=2 | 50 | 1.061(0.876,0.878) | 1.061(0.958,0.942) | 1.069(0.978,0.960) |
| | 100 | 1.049(0.910,0.904) | 1.046(0.958,0.948) | 1.048(0.986,0.974) |
| | 200 | 1.047(0.878,0.876) | 1.045(0.940,0.938) | 1.051(0.980,0.982) |
| | 300 | 1.044(0.896,0.890) | 1.047(0.934,0.934) | 1.047(0.964,0.974) |
| $\alpha$=3 | 50 | 1.059(0.896,0.858) | 1.058(0.958,0.936) | 1.060(0.964,0.962) |
| | 100 | 1.048(0.914,0.914) | 1.052(0.940,0.934) | 1.050(0.972,0.966) |
| | 200 | 1.045(0.894,0.894) | 1.048(0.950,0.936) | 1.046(0.980,0.980) |
| | 300 | 1.045(0.914,0.904) | 1.044(0.974,0.970) | 1.044(0.978,0.976) |
| $\alpha$=5 | 50 | 1.052(0.908,0.900) | 1.059(0.956,0.944) | 1.062(0.976,0.962) |
| | 100 | 1.056(0.924,0.906) | 1.057(0.946,0.940) | 1.055(0.982,0.968) |
| | 200 | 1.049(0.874,0.866) | 1.044(0.938,0.922) | 1.046(0.986,0.974) |
| | 300 | 1.004(0.882,0.888) | 1.045(0.950,0.952) | 1.050(0.990,0.990) |

Table 3.9: $AI_9$ vs $AI_1$
$\hat{r}_9\,(\widehat{CP}_{AI_9}, \widehat{CP}_{AI_1})$

higher empirical finite-sample coverage probabilities as compared to $AI_1$, with the difference $\widehat{CP}_{AI_9} - \widehat{CP}_{AI_1}$ ranging from $-1\%$ to $3.6\%$.

As to analyzing $AI_{10}$ versus $AI_1$ in Table 3.10, while $AI_{10}$ is mostly a bit longer than $AI_1$ on average, the ratio $\hat{r}_{10}$ is less than 1 for $n = 50$, and it increases as $n$ increases and decreases as the confidence level increases. Further, the difference $\widehat{CP}_{AI_{10}} - \widehat{CP}_{AI_1}$ of the empirical finite-sample coverage probabilities of $AI_{10}$ and $AI_1$ fluctuates between $-1.6\%$ and $5.2\%$ and is mostly positive. It turns out that $AI_{10}$ is one of the best performing asymptotic CI's

| | | $1-\gamma$ | | |
|---|---|---|---|---|
| Pareto(I)($\alpha$,1) | $n$ | 0.9 | 0.95 | 0.98 |
| $\alpha$=0.5 | 50 | 1.002(0.920,0.868) | 0.974(0.944,0.930) | 0.930(0.970,0.962) |
| | 100 | 1.013(0.902,0.906) | 1.008(0.950,0.934) | 0.989(0.980,0.976) |
| | 200 | 1.033(0.904,0.900) | 1.019(0.960,0.960) | 1.016(0.980,0.980) |
| | 300 | 1.036(0.900,0.906) | 1.029(0.942,0.934) | 1.027(0.974,0.970) |
| $\alpha$=1.5 | 50 | 0.998(0.912,0.888) | 0.972(0.932,0.928) | 0.933(0.974,0.966) |
| | 100 | 1.021(0.876,0.874) | 1.010(0.952,0.926) | 0.994(0.976,0.964) |
| | 200 | 1.039(0.904,0.902) | 1.028(0.958,0.946) | 1.019(0.978,0.976) |
| | 300 | 1.037(0.880,0.870) | 1.030(0.960,0.952) | 1.027(0.988,0.982) |
| $\alpha$=2 | 50 | 0.998(0.890,0.878) | 0.972(0.948,0.942) | 0.943(0.982,0.960) |
| | 100 | 1.018(0.922,0.904) | 1.003(0.960,0.948) | 0.986(0.974,0.974) |
| | 200 | 1.031(0.878,0.876) | 1.024(0.942,0.938) | 1.020(0.984,0.982) |
| | 300 | 1.034(0.902,0.890) | 1.033(0.946,0.934) | 1.026(0.982,0.974) |
| $\alpha$=3 | 50 | 0.996(0.902,0.858) | 0.970(0.962,0.936) | 0.935(0.974,0.962) |
| | 100 | 1.017(0.914,0.914) | 1.008(0.952,0.934) | 0.988(0.964,0.966) |
| | 200 | 1.029(0.894,0.894) | 1.026(0.950,0.936) | 1.015(0.988,0.980) |
| | 300 | 1.035(0.912,0.904) | 1.030(0.962,0.970) | 1.024(0.980,0.976) |
| $\alpha$=5 | 50 | 0.990(0.922,0.900) | 0.971(0.928,0.944) | 0.937(0.970,0.962) |
| | 100 | 1.025(0.920,0.906) | 1.013(0.954,0.940) | 0.993(0.992,0.968) |
| | 200 | 1.034(0.862,0.866) | 1.023(0.942,0.922) | 1.015(0.982,0.974) |
| | 300 | 1.038(0.878,0.888) | 1.030(0.950,0.952) | 1.029(0.982,0.990) |

Table 3.10: $AI_{10}$ vs $AI_1$
$\hat{r}_{10}\,(\widehat{CP}_{AI_{10}}, \widehat{CP}_{AI_1})$

for $\alpha$ among $AI_1$-$AI_{10}$, as will be discussed in Chapter 4.

We note in passing that using the Computational Cluster for the Department of Statistics at the University of Manitoba (which is four Dell PowerEdge R620 servers, each with a total of 32 CPU cores and 64 GB of memory)[1] to generate Table 3.9 and 3.10, it took 0.347, 5.542, 91.543 and 455.194 seconds to calculate $\hat{\alpha}_{gme}$ of (3.60) for one repetition of the Pareto(I)($\alpha$,1) sample of size $n = 50$,

[1]Each system in the cluster scores over 272 GFlops on the 2017 Intel(R) Optimized LINPACK Benchmark.

100, 200 and 300, respectively. The runtime seems to increase exponentially as $n$ goes up. It would take approximately 50 days to complete each of Tables 3.9 and 3.10 entry-by-entry, but we splited our tables by entries and run the respective simulations as parallel files to improve on time. Then, the obtained results were combined into Tables 3.9 and 3.10 manually rather than by using the function **makeTable** in section A.4 of Appendix which was used to make Tables 3.2 - 3.8.

## 3.4 Scale parameter $\sigma$ is unknown

All the asymptotic CI's for $\alpha$ in sections 3.1-3.3 were built under the assumption that the scale parameter $\sigma$ of the Pareto(I)($\alpha$,$\sigma$) distribution is known. Now let's consider the case when $\sigma$ is unknown and briefly discuss if some of the methods used in sections 3.1-3.3 can be adapted to construct asymptotic CI's for $\alpha$ in this case as well.

First, we consider the asymptotic normality of the Student $t$-statistic as in (3.11) that is based on Pareto(I)($\alpha, \sigma$). To adapt this convergence in distribution to the case when $\sigma$ is unknown, we would first need to replace $\sigma$ in the mean $\alpha\sigma/(\alpha - 1)$ with its consistent estimator that, for example, converges in probability to $\sigma$ at an appropriate rate or is asymptotically normal with an appropriate rate, and then perform further detailed analysis to see if such a modified left-hand side of (3.11) would converge in distribution as $n \to \infty$. This investigation is beyond the scope of this thesis.

Next, we look at the convergence of (3.16), the counterpart of (3.11) with the $Exp(\alpha)$ i.i.d. r.v.'s $\{Y_i\}_{1\le i\le n} = \{log\,(X_i/\sigma)\}_{1\le i\le n}$ (cf. (2.3)) that was used to build $AI_1$, where $X_1$,...,$X_n$ is a random sample from Pareto(I)($\alpha$,$\sigma$). Clearly, unless $\sigma$ is known, the key relationship (2.3) no longer holds true. For this reason, we cannot adapt (3.16) and also the FCLT based CI's obtained in section 3.2 here. However, the convergences in section 3.2 can be applied to $X_1$, $X_2$,... directly (provided that $\alpha \ge 2$) and then potentially modified when $\sigma$ can be appropriately estimated there, similarly to the discussions in the previous paragraph.

Now, let's look at the $\tilde{\alpha}_{mle}$ of (2.15), which is the MLE of $\alpha$ when $\sigma$ is unknown. From (2.17) and the CLT, we get

$$\frac{\dfrac{2\alpha n}{\tilde{\alpha}_{mle}} - (2n-2)}{2\sqrt{n-1}} \xrightarrow[n\to\infty]{D} N(0,1),$$

which leads to the following $(1-\gamma)100\%$ asymptotic CI for $\alpha$:

$$\left[\tilde{\alpha}_{mle}\left(\frac{n-1}{n} - \frac{z_{\frac{\gamma}{2}}\sqrt{n-1}}{n}\right), \tilde{\alpha}_{mle}\left(\frac{n-1}{n} + \frac{z_{\frac{\gamma}{2}}\sqrt{n-1}}{n}\right)\right].$$

In Quandt (1966), the MME estimators $\tilde{\alpha}_{mme}$ and $\tilde{\sigma}_{mme}$ for $\alpha$ and $\sigma$ when both parameters are unknown were obtained by equating the sample mean and sample minimum to their expectations,

$$\tilde{\alpha}_{mme} = \frac{n\bar{X}_n - X_{1:n}}{\bar{X}_n - X_{1:n}} \qquad \text{and} \qquad \tilde{\sigma}_{mme} = (1-(n\tilde{\alpha}_{mme})^{-1})X_{1:n}.$$

Both of these estimators were found to be strongly consistent. However, we have not found any literature on whether $\tilde{\alpha}_{mme}$, properly centered and scaled, has an asymptotic distribution so that this could potentially be used to construct an asymptotic CI for $\alpha$. As $\tilde{\alpha}_{mme}$ is a function of $(\bar{X}_n, X_{1:n})$, as long as the limiting distribution of the latter vector exists and can be found, one can hope to find the limiting distribution of $\tilde{\alpha}_{mme}$. Such asymptotic studies are beyond the scope of this thesis.

One can also think of another set of the MME estimators for $\alpha$ and $\sigma$ by equating the sample mean and variance to the mean and variance of Pareto(I)($\alpha$,$\sigma$):

$$\bar{X}_n = \frac{\alpha\sigma}{\alpha - 1} \quad \text{and} \quad \overline{X^2}_n - (\bar{X}_n)^2 = \frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}.$$

The corressponidng MME $\tilde{\alpha^*}_{mme}$ for $\alpha$ is a solution of the quadratic equation

$$\alpha^2 - 2\alpha - \frac{(\bar{X}_n)^2}{\overline{X^2}_n - (\bar{X}_n)^2} = 0.$$

Since $\tilde{\alpha^*}_{mme}$ is a function of $(\bar{X}_n, \overline{X^2}_n)$, one can first try to establish asymptotic normality of the latter vector, properly centered and normalized, and then proceed with finding the limiting distribution of $\tilde{\alpha}^*_{mme}$, as $n \to \infty$, potentially leading to an asymptotic CI for $\alpha$.

Finally, in Brazauskas and Serfling (2000b), the authors presented the GME for $\alpha$ when $\sigma$ is unknown as follows. First, the kernel $h_0^*(\cdot)$ was defined as

$$h_0^*(X_1, ..., X_k) = \frac{k}{\sum_{i=1}^k (log X_i - log \min_{1\le i\le k}(X_1, ..., X_k))}, \qquad (3.64)$$

which is the MLE for $\alpha$ as in (2.15) based on the sample $X_1$,...,$X_k$ of size $k$. Similarly to the construction of $\hat{\alpha}_{gme}$ in subsection 3.3.3,

$$\frac{2k\alpha}{h_0^*(X_1, ..., X_k)} \sim \chi^2_{2n-2}. \qquad (3.65)$$

The latter relationship can be used to see that the median of the r.v.

$$h^*(X_1, ..., X_k) = \frac{M_{2k-2}}{2k} h_0^*(X_1, ..., X_k) \qquad (3.66)$$

becomes $\alpha$, where $M_{2k-2}$ is the median of $\chi^2_{2k-2}$. Finally, similarly to $\hat{\alpha}_{gme}$ of (3.60), they defined $\hat{\alpha}^*_{gme}$ of $\alpha$ as the median of all $h^*(\cdot)$'s based on all $\binom{n}{k}$ subsamples of the sample $X_1, ..., X_n$:

$$\hat{\alpha}^*_{gme} = Median\left\{h^*(X_{i_1}, ..., X_{i_k}), 1 \le i_1 < ... < i_k \le n\right\}.$$

Brazauskas and Serfling (2000b) proved that (3.61) holds true with $\hat{\alpha}^*_{gme}$ and a constant $\gamma^*_k$ replacing $\hat{\alpha}_{gme}$ and $\gamma_k$, and tabulated $\gamma^*_k$ for $k$=2:10. Consequently, we can use $AI_9$ and $AI_{10}$ with $\hat{\alpha}^*_{gme}$ and $\gamma^*_k$ instead of $\hat{\alpha}_{gme}$ and $\gamma_k$ as the $100(1-\gamma)\%$ asymptotic CI's for $\alpha$ when $\sigma$ is unknown.

# Chapter 4

# Conclusions

In sections 2.1 and 3.1 - 3.3, we derived various exact and asymptotic CI's for the tail index $\alpha$ of Pareto(I)($\alpha$,$\sigma$) using several pivotal quantities and assuming that the scale parameter $\sigma$ is known. In sections 2.2 and 3.4, we briefly outlined a few exact and asymptotic CI's for $\alpha$ obtained without the latter assumption. We minimized the expected lengths of a shorter of the two exact CI's built in section 2.1 and its analogue in section 2.2. Asymptotic CI's $AI_2$-$AI_{10}$ constructed in sections 3.1 - 3.3 were compared one-by-one to the CLT based one in (3.18), $AI_1$, in terms of their expected lengths and finite-sample coverage probabilites. Our CLT and FCLT based asymptotic CI's of sections 3.1 and 3.2 were built on the $Exp(\alpha)$ transformation of our original Pareto(I)($\alpha$,$\sigma$) sample as in (2.3). This way we avoided imposing the condition $\alpha \geq 2$ (implying that Pareto(I)($\alpha$,$\sigma$) $\in$ DAN and allowing the use of those CLT and FCLT for the Pareto(I)($\alpha$,$\sigma$) distribution). It also led to the CI's with shorter expected lengths and higher finite-sample coverage probabilities.

In this chapter, we will first present a summary table of the performance of our asymptotic CI's $AI_2$-$AI_{10}$ versus that of $AI_1$ (cf. Table 4.1), and then determine the better ones among them, that is those with relatively short expected lengths and appropriately high finite-sample coverage probabilities. For a convenient overall comparison of $AI_1$-$AI_{10}$, Table 4.1 reports the ranges of the differences $\widehat{CP}_{AI_k}$-$\widehat{CP}_{AI_1}$ of the empirical finite-sample coverage probabilities of $AI_k$ and $AI_1$ (in %) and the average values of the ratios $\hat{r}_k$ of the empirical expected lengths of $AI_k$ and $AI_1$ taken over different values of the tail index $\alpha$ and the confidence level $1-\gamma$ from Tables 3.2 - 3.10, $k = \overline{2, 10}$.

From Tables 3.2 - 3.10 and Table 4.1, we conclude that the empirical finite-sample coverage probabilities of $AI_2$-$AI_{10}$ are higher than those of $AI_1$, except for a few cases for some of $AI_2$-$AI_{10}$ where $AI_1$ has a slightly higher coverage. As to the expected lengths, the CI $AI_7$ of (3.51), based on the asymptotic normality of the MLE of $\alpha$, is somewhat shorter than $AI_1$ on average, though less so as $n$ increases. The same holds true for the GME based $AI_{10}$ of (3.63) when $n = 50$. We also note that the ratios $\hat{r}_7$ and $\hat{r}_{10}$ of the expected lengths of $AI_7$ and $AI_{10}$ to that of $AI_1$ decrease as the confidence level $1-\gamma$ increases (cf. Table 3.7 and 3.10). The rest of our asymptotic CI's for $\alpha$ are longer than $AI_1$ on average. Consequently, $AI_7$ performs better than $AI_1$ both in terms of the expected length and finite-sample coverage probability and hence may be an appealing choice among $AI_1$-$AI_{10}$. It overall improves on another MLE based asymptotic CI, $AI_6$ of (3.48), that has a slightly higher finite-sample

| Asymptotic CI's | | $n$ 50 | 100 | 200 | 300 | 1000 |
|---|---|---|---|---|---|---|
| $AI_2$ | Range of $\widehat{CP}_{AI_2}$-$\widehat{CP}_{AI_1}$ | [-0.3, 0.4] | [-0.1, 0.4] | - | [-0.2, 0.3] | [-0.2, 0.2] |
| | Mean of $\hat{r}_2$ | 1.075 | 1.063 | - | 1.057 | 1.055 |
| $AI_3$ | Range of $\widehat{CP}_{AI_3}$-$\widehat{CP}_{AI_1}$ | [0.5, 2] | [0.2, 1.5] | - | [0.1, 1.1] | [-0.1, 0.8] |
| | Mean of $\hat{r}_3$ | 1.102 | 1.075 | - | 1.054 | 1.042 |
| $AI_4$ | Range of $\widehat{CP}_{AI_4}$-$\widehat{CP}_{AI_1}$ | [0.5, 1.3] | [-0.2, 0.7] | - | [-0.3, 0.6] | [-0.3, 0.2] |
| | Mean of $\hat{r}_4$ | 1.258 | 1.199 | - | 1.168 | 1.159 |
| $AI_5$ | Range of $\widehat{CP}_{AI_5}$-$\widehat{CP}_{AI_1}$ | [0.3, 1] | [-0.3, 0.5] | - | [-0.2, 0.7] | [-0.4, 0.3] |
| | Mean of $\hat{r}_5$ | 1.166 | 1.13 | - | 1.11 | 1.104 |
| $AI_6$ | Range of $\widehat{CP}_{AI_6}$-$\widehat{CP}_{AI_1}$ | [0.6, 1.4] | [0.3, 1.1] | - | [0.1, 0.6] | [-0.1, 0.4] |
| | Mean of $\hat{r}_6$ | 1.017 | 1.009 | - | 1.004 | 1.001 |
| $AI_7$ | Range of $\widehat{CP}_{AI_7}$-$\widehat{CP}_{AI_1}$ | [1.7, 3] | [0.8, 1.7] | - | [0.2, 0.8] | [0, 0.4] |
| | Mean of $\hat{r}_7$ | 0.936 | 0.969 | - | 0.99 | 0.997 |
| $AI_8$ | Range of $\widehat{CP}_{AI_8}$-$\widehat{CP}_{AI_1}$ | [2.2, 8.3] | [1.3, 4.8] | - | [0.3, 2.4] | [-0.1, 1] |
| ($\alpha = 3 \,\&\, 5$) | Mean of $\hat{r}_8$ | 1.041 | 1.069 | - | 1.086 | 1.0915 |
| $AI_9$ | Range of $\widehat{CP}_{AI_9}$-$\widehat{CP}_{AI_1}$ | [-0.2, 3.8] | [-0.4, 1.8] | [-0.6, 1.6] | [-1, 1] | - |
| | Mean of $\hat{r}_9$ | 1.06 | 1.051 | 1.047 | 1.043 | - |
| $AI_{10}$ | Range of $\widehat{CP}_{AI_{10}}$-$\widehat{CP}_{AI_1}$ | [-1.6, 5.2] | [-0.4, 2.6] | [-0.4, 2] | [-1, 1.2] | - |
| | Mean of $\hat{r}_{10}$ | 0.968 | 1.006 | 1.025 | 1.031 | - |

Table 4.1: Summary performance of asymptotic CI's $AI_2$-$AI_{10}$ vs that of $AI_1$ Ranges of $\widehat{CP}_{AI_k}$-$\widehat{CP}_{AI_1}$ (in %) and means of $\hat{r}_k$ for $AI_2$-$AI_{10}$

coverage probability than that of $AI_1$ and is insignificantly longer than $AI_1$.

On the other hand, while the empirical expected lengths of $AI_{10}$ are under, or slightly above those of $AI_1$ (cf. Table 4.1), the empirical finite-sample coverage probabilities of $AI_{10}$ are the highest among all $AI_1$-$AI_{10}$, excluding the MME based $AI_8$ derived under the restriction $\alpha > 2$ and examined for $\alpha = 3$ and 5 only. Therefore, $AI_{10}$ may be desirable when a CI for $\alpha$ with a higher finite-sample coverage probability is a priority. For practical applications, one may like to improve on time efficiency of computing $AI_{10}$ along the lines proposed in Brazauskas and Serfling (2000a). We recall that, due to a computational cost, the simulations for $AI_{10}$ in Table 3.10 (and for $AI_9$ in Table 3.9) were done for the sample sizes $n = 50$, 100, 200, and 300 and the number of repetitions $N = 500$ instead of for $n = 50$, 100, 300, and 1000 and $N = 10{,}000$ used for $AI_1$-$AI_8$. From Table 4.1, we also observe that $AI_{10}$ outperforms $AI_9$ of (3.62), another asymptotic GME based CI for $\alpha$ that may also be desirable in practice due to its probability coverage and relatively short expected length.

Consider the MME based $AI_8$ of (3.56) derived under the restriction that $\alpha > 2$. Its empricial finite-sample coverage probability outperforms that of $AI_1$ by the highest percentage of 8.3% for $n = 50$ and Pareto(I)(3,1) (cf. Table 3.8), but at the expense of $AI_8$ being longer than $AI_1$ on average. For the case of Pareto(I)(5,1), from Table 3.8, the expected length of $AI_8$ is under, or slightly above, that of $AI_1$, and the coverage $\widehat{CP}_{AI_8}$ is almost at, or above, the nominal $1 - \gamma$ and is higher than $\widehat{CP}_{AI_1}$.

Analyzing the rest of the asymptotic CI's in Table 4.1 in the light of an apparent trade-off between their finite-sample coverage probabilities and expected lengths, we observe that $AI_3$ of (3.33) has the highest coverage among the FCLT based CI's, up to 2% higher than that of $AI_1$, and its expected length is relatively short.

Since the finite-sample coverage probabilities of $AI_7$ of (3.51) and $AI_{10}$ of (3.63) are close enough to the nominal $1-\gamma$ level (cf. Tables 3.7 and 3.10), and since they perform well in terms of their relative expected lengths as well, it seems natural to compare the expected lengths of $AI_7$ and $AI_{10}$ with that of $I_1$ of (2.7), the better performing of the two exact CI's from section 2.1. First, from (2.9) and (3.51),

$$\frac{E(\text{Length of } AI_7)}{E(\text{Length of } I_1)} = \frac{\frac{2\alpha n}{n-1}\frac{z_{\frac{\gamma}{2}}}{\sqrt{n}}}{\frac{\alpha}{2(n-1)}(\chi^2_{2n,1-\gamma_2} - \chi^2_{2n,\gamma_1})}$$

$$= \frac{4z_{\frac{\gamma}{2}}\sqrt{n}}{\chi^2_{2n,1-\gamma_2} - \chi^2_{2n,\gamma_1}},$$

where the quantiles $\chi^2_{2n,1-\gamma_2}$ and $\chi^2_{2n,\gamma_1}$ of the $\chi^2_{2n}$ distribution (cf. (2.5)) that minimize $E(\text{Length of } I_1)$ are taken from Table 2.1. By examing this ratio numerically for our values of $n = 50$, 100, 300, and 1,000 and $1-\gamma = 0.9$, 0.95, and 0.98, we conclude that $AI_7$ is only insignificantly longer than $I_1$ on average, by a factor of less than 1.007 for $n = 50$ and even less so for larger values of $n$. A similar conclusion holds true for $\widehat{E(\text{Length of } AI_{10})}/\widehat{E(\text{Length of } I_1)}$

that seems to slightly deviate around 1.04 for $\alpha = 0.5$, 1.5, 2, 3, and 5, $n = 50$, 100, 200, and 300, and $1 - \gamma = 0.9$, 0.95, and 0.98. However, the latter ratio can be seen to go a bit below one (and above 0.99) for $n = 30$ when $AI_{10}$ of (3.63) is defined with $k = 10$ instead of $k = 4$, in which case the asymptotic relative efficiency (with respect to the MLE of $\alpha$) of $\hat{\alpha}_{gme}$ is known to be higher (cf. Brazauskas and Serfling (2000a)). Thus, although the performance of the exact CI $I_1$ and those of the asymptotic CI's $AI_7$ and $AI_{10}$ are very comparable in terms of the finite-sample coverage probabilities and expected lengths, it may not be the case when the actual data departures from the assumed Pareto(I)($\alpha$,$\sigma$) distribution, for example due to unrepresentative outliers in the sample. Moreover, in view of robustness of $\hat{\alpha}_{gme}$ (cf. Brazauskas and Serfling (2000a) and subsection 3.3.3 in this thesis), we conjecture that in the case of such a deviation, $AI_{10}$ likely outperforms $I_1$. Perhaps, this is also true for $AI_7$ and some other asymptotic CI's for $\alpha$ in this thesis.

Finally, in addition to investigating the veracity of our conjecture at the end of the previous paragraph, other possible extensions of the research of this thesis could include studying CI's for $\alpha$ obtained from other possible pivotal quantities, for example by considering other convenient functionals of the Student process in the FCLT of (3.20). It would also be of interest to investigate the CI's for $\alpha$ outlined in sections 2.2 and 3.4 when the parameter $\sigma$ is unknown. Constructing exact and asymptotic CI's for $\sigma$ and confidence regions for ($\alpha$,$\sigma$) when both $\alpha$ and $\sigma$ are unknown could be another natural desirable extension of our work.

# Appendix

# R code

## A.1 Finding the quantiles of a chi-square distribution that minimize the length of $I_1$

**find.cutoffs** Given $n$ (the sample size) and $\gamma$ (the value of 1-confidence level), the function **find.cutoffs** finds the upper and lower quantiles $a$ and $b$ of $\chi^2_{2n}$ as in (2.10) which minimize the length of the exact confidence interval $I_1$ of (2.7). To find such $a$ and $b$, the code is implemented to follow the algorithm described on page 10 in this thesis.

**find.b** Given values of $n$ and $a$, the function **find.b** finds the value of $b$ that satisfies $g(a)$=$g(b)$ as in (2.10).

```
###############################################################################
#find.cutoffs
###############################################################################
find.cutoffs<-function(args){
  n<-args[1]
  g<-args[2]
  g<-1-g
  chi.mode<-max(c(n-2,0))
  leftratio<-0.45
  rightratio<-0.999999
  a.left<-leftratio*chi.mode
```

```
  b.left<-find.b(a.left,n,chi.mode)
  cov.left<-pchisq(b.left,n)-pchisq(a.left,n)
  a.right<-rightratio*chi.mode
  b.right<-find.b(a.right,n,chi.mode)
  cov.right<-pchisq(b.right,n)-pchisq(a.right,n)

  if(abs(cov.left-g)<10^-7){
    a<-a.left
    b<-b.left
    cov<-cov.left
    return(c(a,b,cov))
  }
  else if(abs(cov.right-g)<10^-7){
    a<-a.right
    b<-b.right
    cov<-cov.right
    return(c(a,b,cov))
  }
  else{
    a.cen<-(a.left+a.right)/2
    b.cen<-find.b(a.cen,n,chi.mode)
    cov.cen<-pchisq(b.cen,n)-pchisq(a.cen,n)
  }
  while(abs(cov.cen-g)>10^-7){
    if(cov.cen<g){
      a.right<-a.cen
      b.right<-b.cen

      a.cen<-(a.left+a.right)/2 #finding the new a_M
      b.cen<-find.b(a.cen,n,chi.mode)
      cov.cen<-pchisq(b.cen,n)-pchisq(a.cen,n)
    }
    if(cov.cen>=g){
      a.left<-a.cen
      b.left<-b.cen

      a.cen<-(a.left+a.right)/2 #finding the new a_M
      b.cen<-find.b(a.cen,n,chi.mode)
      cov.cen<-pchisq(b.cen,n)-pchisq(a.cen,n)
    }
  }
  return(paste("(",formatC(a.cen,format="f",digits=3),",",formatC(b.cen,
      format="f",digits=3),")",sep=""))
}
###############################################################################
#find.b
###############################################################################
find.b<-function(a,n,chi.mode){
  a.val<-dchisq(a,n)
  upper<-3*chi.mode
  b<-uniroot(function(x) a.val-dchisq(x,n),lower=chi.mode,upper=upper,tol=1e
      -9)$root
  return(b)
}
```

## A.2 Building Tables 2.1 and 2.2 of Chapter 2

The code below builds Tables 2.1 and 2.2, which are the tables of the quantiles $(a, b)$ of $\chi^2_{2n}$ as in (2.10) and the quantiles $(\chi^2_{2n-2,a}, \chi^2_{2n-2,b})$ of $\chi^2_{2n-2}$ that minimize the length of the CI in (2.19).

```
##############################################################################
#Table 2.1
##############################################################################
g.s<-c(0.1,0.05,0.02)
g.vec<-c(rep(g.s[1],4),rep(g.s[2],4),rep(g.s[3],4))
n.values=c(50,100,300,1000)
df.values<-n.values*2
df.vec=rep(df.values,3)
args=matrix(c(df.vec,g.vec),12,2)
cutoffs.vec<-apply(args,1,find.cutoffs)
cutoffs.mat<-matrix(cutoffs.vec,4,3)
rownames(cutoffs.mat)<-n.values
colnames(cutoffs.mat)<-1-g.s
View(cutoffs.mat)
##############################################################################
#Table 2.2
##############################################################################
g.s<-c(0.1,0.05,0.02)
g.vec<-c(rep(g.s[1],4),rep(g.s[2],4),rep(g.s[3],4))
n.values=c(50,100,300,1000)
df.values.2<-n.values*2-2
df.vec.2=rep(df.values.2,3)
args2=matrix(c(df.vec.2,g.vec),12,2)
cutoffs.vec2<-apply(args2,1,find.cutoffs)
cutoffs.mat2<-matrix(cutoffs.vec2,4,3)
rownames(cutoffs.mat2)<-n.values
colnames(cutoffs.mat2)<-1-g.s
View(cutoffs.mat2)
```

## A.3 Generating asymptotic CI's of Chapter 3

**AI1** Given a random sample $x$ of size $n$ from Pareto(I)($\alpha$,$\sigma$) and g (the value of 1-confidence level), the function **AI1** builds $AI_1$ of (3.18). Additionally, the true $\alpha$ is given as a parameter to check if the resulting interval contains it. The scale parameter $\sigma$ is also passed as a parameter, though we use $\sigma$=1. As a result, the function returns the length of $AI_1$ and a binary number indicating whether $AI_1$ captures the true $\alpha$, where 1 refers to $AI_1$ capturing the true $\alpha$. In the case when a sample does not satisfy the condition that the lower bound in (3.17) is positive, the code stops.

**AI2-AI10** Each of the functions **AI2** to **AI10** for the CI's $AI_2$ to $AI_{10}$ takes a desired number $n$ of the sample size and g as parameters, generates a random sample of size $n$ from Pareto(I)($\alpha$,$\sigma$) and builds the corresponding $1 - \text{g}$ size CI. As in **AI1**, the true $\alpha$ is passed as a parameter to check if the resulting intervals contain it. The value of $\sigma$ is also a parameter, though we take $\sigma = 1$. Within each of these functions **AI2-AI10**, the function **AI1** is called with the same sample generated by these functions for comparison of the corresponding interval with $AI_1$. As a result, the lengths of the latter two CI's and whether the CI's capture the true $\alpha$ are returned. Note that each of $AI_2$-$AI_6$, $AI_8$ and $AI_9$ was derived under certain conditions, as described in Chapter 3. In the case when the initial sample generated by the functions **AI2-AI6**, **AI8** and **AI9** does not satisfy the conditions of the validity of their corresponding CI's, these functions ignore such a "bad" sample, resample and repeat the process until the new sample satisfies the conditions. However, **AI9** is built to stop the code with an error message if a "bad" sample arises. An additional parameter for the functions **AI2-AI8**, add.condition, which is set equal to TRUE by default for the purposes of the present thesis, enables us to use only the samples that satisfy the conditions of the validity of *all* $AI_1$-$AI_6$ and $AI_8$ (that are checked via the function **condition** listed next). This parameter can be set equal to FALSE, if one does not need such samples. Note that add.condition does not count in the condition for the validity of $AI_9$, due to the computational cost of this CI. The functions **AI2-AI10** are inspired by Tuzov (2014).

**condition** Given a random sample $x$ from Pareto(I)($\alpha$,$\sigma$), its exponential counterpart $y$ as in (3.15), g (the value of 1-confidence level), and the values of $\alpha$ and $\sigma$, the function **condition** checks if this particular sample $x$ satisfies the conditions for the validity (derivation) of the CI's $AI_1$-$AI_6$ and $AI_8$ and returns FALSE only when it satisfies *all* these conditions. The condition for deriving $AI_9$ is not included due to its computational time.

**GME** The function GME takes a random sample $x$ of size $n$ from Pareto(I)($\alpha$,$\sigma$) and $k$ (for $\binom{n}{k}$ subsamples, $k = 4$ for Tables 3.9 and 3.10) and returns the GME for $\alpha$ as in (3.60).

```
##########################################################################
#AI1
##########################################################################
AI1<-function(x,n,g,alpha,sigma){
  y<-log(x/sigma)
  lower<-n/(sum(y)+qnorm(1-g/2,0,1)*sd(y)*sqrt(n))
  upper<-n/(sum(y)-qnorm(1-g/2,0,1)*sd(y)*sqrt(n))
  length<-upper-lower

  if((1/upper)<0) stop("The sample does not satisfy the condition")
  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in))
}
##########################################################################
#AI2
##########################################################################
AI2<-function(n,g,alpha,sigma,add.condition=TRUE){
  cond<--1
  t0<-0.9
  while( cond<0 ||add.condition){
    x<-rpareto(n,alpha,sigma)
    y<-log(x/sigma)
    t0<-0.9
    flor<-floor(n*t0)
    if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
    if((sum(y[1:flor])+qnorm(1-g/2,0,1)*sd(y)*sqrt(flor)/flor)>0) cond<-1
  }
  lower<-flor/(sum(y[1:flor])+qnorm(1-g/2,0,1)*sd(y)*sqrt(flor))
  upper<-flor/(sum(y[1:flor])-qnorm(1-g/2,0,1)*sd(y)*sqrt(flor))
  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
##########################################################################
#AI3
##########################################################################
AI3<-function(n,g,alpha,sigma,add.condition=TRUE){
  b.s<-c(1.96,2.241538,2.576)
  if(g==0.1){b=b.s[1]}
```

```
  if(g==0.05){b=b.s[2]}
  if(g==0.02){b=b.s[3]}

  checker<--1
  while(checker<0||add.condition){
    x<-rpareto(n,alpha,sigma)
    y<-log(x/sigma)
    k<-1:n
    check<-(cumsum(y)-b*sd(y)*sqrt(n))/k
    neg.indices<-which(check<0)
    if(length(neg.indices)!=n) checker=1
    if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
  }
  lower.part<-(cumsum(y)+b*sd(y)*sqrt(n))/k
  upper.part<-(cumsum(y)-b*sd(y)*sqrt(n))/k

  lower.part<-1/lower.part
  upper.part<-1/upper.part
  upper.part<-upper.part[-neg.indices]

  lower<-max(lower.part)
  upper<-min(upper.part)

  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
################################################################################
#AI4
################################################################################
AI4<-function(n,g,alpha,sigma,add.condition=TRUE){
  cond<--1
  while(cond<0||add.condition){
    x<-rpareto(n,alpha,sigma)
    y<-log(x/sigma)
    sum.sum.x<-sum(cumsum(y[1:(n-1)]))
    if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
    if(((sum(cumsum(y[1:(n-1)]))/n)-(qnorm(1-g/2,0,1)*sd(y)*sqrt(n)/sqrt(3)))
        /((n-1)/2)>0) cond<-1
  }
  lower<-((n-1)/2)/((sum(cumsum(y[1:(n-1)]))/n)+(qnorm(1-g/2,0,1)*sd(y)*sqrt(
      n)/sqrt(3)))
  upper<-((n-1)/2)/((sum(cumsum(y[1:(n-1)]))/n)-(qnorm(1-g/2,0,1)*sd(y)*sqrt(
      n)/sqrt(3)))

  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
################################################################################
#AI5
################################################################################
AI5<-function(n,g,alpha,sigma,add.condition=TRUE){
  b.s<-c(0.765,0.94,1.195,1.655,2.29,2.79)
```

```
  gs<-c(0.2,0.15,0.1,0.05,0.02,0.01)
  values<-matrix(c(gs,b.s),length(gs),2,byrow=F)
  b=values[values[,1]==g,2] # select a value of b according to alpha
  cond<--1
  while(cond<0||add.condition){
    x<-rpareto(n,alpha,sigma)
    y<-log(x/sigma)
    k<-seq(1,n-1)
    sum.x<-cumsum(y[1:(n-1)])
    sum.x.k<-sum(k*sum.x)
    sum.x.sq<-sum((sum.x)^2)
    c<-(n*(n-1)*(2*n-1))/6
    Sn<-sd(y)
    Dn<-sum.x.k^2-c*(sum.x.sq-b*(Sn*n)^2)
    if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
    if(Dn>0&&sum.x.sq-b*(Sn*n)^2>0) cond<-1
  }
  Dn<-sqrt(Dn)
  lower<-c/(sum.x.k+Dn)
  upper<-c/(sum.x.k-Dn)

  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
##############################################################################
#AI6
##############################################################################
AI6<-function(n,g,alpha,sigma,add.condition=TRUE){
  cond<--1
  while(cond<0||add.condition){
    x<-rpareto(n,alpha,sigma)
    y<-log(x/sigma)
    mle<-a.mle(x,sigma)
    if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
    if((1-(qnorm(1-g/2,0,1)/sqrt(n)))/mle>0) cond<-1
  }
  lower<-mle/(1+(qnorm(1-g/2,0,1)/sqrt(n)))
  upper<-mle/(1-(qnorm(1-g/2,0,1)/sqrt(n)))
  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
##############################################################################
#AI7
##############################################################################
AI7<-function(n,g,alpha,sigma,add.condition=TRUE){
  cond<--1
  while(cond<0||add.condition){
    x<-rpareto(n,alpha,sigma)
    y<-log(x/sigma)
    mle<-a.mle(x,sigma)
    if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
    cond<-1
```

```
  }
  lower<-mle-(mle*qnorm(1-g/2,0,1))/sqrt(n)
  upper<-mle+(mle*qnorm(1-g/2,0,1))/sqrt(n)
  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
##############################################################################
#AI8
##############################################################################
AI8<-function(n,g,alpha,sigma,add.condition=TRUE){
  if(alpha<=2){
    check<--1
    while(check<0||add.condition){
      x<-rpareto(n,alpha,sigma)
      y<-log(x/sigma)
      add.condition<-condition(x,y,g,alpha,sigma)
      if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
      if(alpha<=2) check<-1
    }
    return(c(0,0,0,0))
  }
  else{
    cond<--1
    while( cond<0 ||add.condition){
      x<-rpareto(n,alpha,sigma)
      y<-log(x/sigma)
      mme<-a.mme(x,sigma)
      if(add.condition) add.condition<-condition(x,y,g,alpha,sigma)
      if(mme>2) cond<-1
    }
    z<-qnorm(1-g/2,0,1)
    c<-z*sqrt(mme)*(mme-1)/(sqrt(n*(mme-2)))
    lower.mme<-mme-c
    upper.mme<-mme+c
    length.mme<-upper.mme-lower.mme

    if((lower.mme<=alpha)&&(upper.mme>=alpha)){alpha.in=1} else alpha.in=0
    return(c(length.mme,alpha.in,AI1(x,n,g,alpha,sigma)))
  }
}
##############################################################################
#AI9
##############################################################################
AI9<-function(n,g,alpha,sigma,k=4){
  g.ks<-c(1.563,1.280,1.141,1.088,1.061,1.044,1.035,1.028,1.023,1.019)
  g.k<-g.ks[k]
  x<-rpareto(n,alpha,sigma)
  y<-log(x/sigma)

  a.gm<-GME(x,k)
  lower<-(a.gm*(sqrt(n))/(sqrt(n)+qnorm(1-g/2,0,1)*sqrt(g.k)))
  upper<-(a.gm*(sqrt(n))/(sqrt(n)-qnorm(1-g/2,0,1)*sqrt(g.k)))

  length<-upper-lower
```

```
  if((1/upper)<0) stop("The sample does not satisfy the condition")
  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0

  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
##############################################################################
#AI10
##############################################################################
AI10<-function(n,g,alpha,sigma,k=4){
  g.ks<-c(1.563,1.280,1.141,1.088,1.061,1.044,1.035,1.028,1.023,1.019)
  g.k<-g.ks[k]

  x<-rpareto(n,alpha,sigma)
  y<-log(x/sigma)
  a.gm<-GME(x,k)

  lower<-a.gm-(qnorm(1-g/2,0,1)*a.gm*sqrt(g.k))/sqrt(n)
  upper<-a.gm+(qnorm(1-g/2,0,1)*a.gm*sqrt(g.k))/sqrt(n)

  length<-upper-lower

  if((lower<=alpha)&&(upper>=alpha)){alpha.in=1} else alpha.in=0
  return(c(length,alpha.in,AI1(x,n,g,alpha,sigma)))
}
##############################################################################
#condition
##############################################################################
condition<-function(x,y,g,alpha,sigma){
  n<-length(x)
  mme<-a.mme(x,sigma)
  mle<-a.mle(x,sigma)
  t0<-0.9

  if(((sum(y)-qnorm(1-g/2,0,1)*sd(y)*sqrt(n))/n)>0) cond.1<-TRUE
  else cond.1<-FALSE

  flor<-floor(n*t0)
  if((flor/(sum(y[1:flor])-qnorm(1-g/2,0,1)*sd(y)*sqrt(flor)))>0) cond.2<-
      TRUE
  else cond.2<-FALSE

  k<-1:n
  b.s<-c(1.96,2.241538,2.576)
  if(g==0.1){b=b.s[1]}
  if(g==0.05){b=b.s[2]}
  if(g==0.02){b=b.s[3]}
  check<-(cumsum(y)-b*sd(y)*sqrt(n))/k
  neg.indices<-which(check<0)
  if(length(neg.indices)!=n) cond.3<-TRUE
  else cond.3<-FALSE

  sum.sum.x<-sum(cumsum(y[1:(n-1)]))
  if((((n-1)/2)/((sum(cumsum(y[1:(n-1)]))/n)-(qnorm(1-g/2,0,1)*sd(y)*sqrt(n)/
      sqrt(3))))>0) cond.4<-TRUE
```

```
  else cond.4<-FALSE

  b.s<-c(0.765,0.94,1.195,1.655,2.29,2.79)
  gs<-c(0.2,0.15,0.1,0.05,0.02,0.01)
  values<-matrix(c(gs,b.s),length(gs),2,byrow=F)
  b=values[values[,1]==g,2]
  k<-seq(1,n-1)
  sum.x<-cumsum(y[1:(n-1)])
  sum.x.k<-sum(k*sum.x)
  sum.x.sq<-sum((sum.x)^2)
  c<-(n*(n-1)*(2*n-1))/6
  Dn<-sum.x.k^2-c*(sum.x.sq-b*(sd(y)*n)^2)
  if(Dn>0&&sum.x.sq-b*(sd(y)*n)^2>0) cond.5<-TRUE
  else cond.5<-FALSE

  k<-1:n

  if( (  (1-(qnorm(1-g/2,0,1)/sqrt(n)))/mle)  >0 ) cond.6<-TRUE
  else cond.6<-FALSE

  if(alpha<=2){ cond.8<-TRUE}
  else{
    if(mme>2) cond.8<-TRUE
    else cond.8<-FALSE
  }

  if(cond.1&&cond.2&&cond.3&&cond.4&&cond.5&&cond.6&&cond.8) condition<-TRUE
  else condition<-FALSE
  return(!condition)
}
##############################################################################
#GME
##############################################################################
GME<-function(x,k){
  corrections<-c(0.693,0.839,0.891,0.918,0.934,0.945,0.953,0.959,0.963,0.967)
  correction<-corrections[k]
  n<-length(x)
  comb.mat<-combn(x,k)

  #All combination of size k from a set of sample x of size n.

  comb.mat<-log(comb.mat)
  h0.inv<-colMeans(comb.mat)
  #These two lines build h0.

  h<-correction/h0.inv

  a.gm<-median(h)
  return(a.gm)
}
```

## A.4 Making tables of Chapters 3 and 4

**makeTable** The function **makeTable** is used to make Tables 3.2 to 3.8 and Table 4.1. It takes parameters listed below:

CI: The CI number (from 2 to 8).

$n$: The vector of possible values of $n$ (the sample size).

g: The vector of possible values of $\gamma$ (where $1 - \gamma$ is the confidence level).

$a$: The vector of possible values of the tail index $\alpha$.

sigma: The value of $\sigma$. The default is set to 1.

$N$: The number of repetitions of a Pareto(I)($\alpha$,$\sigma$) sample. The default is set to 10,000.

seed: The seed to be used. The deault is set to 123.

comparison: This binary parameter is set to TRUE by default for the purposes of the present thesis, so that studying $AI_1$-$AI_8$ in Tables 3.2 to 3.8 and the comparative Table 4.1 is based on the same samples that satisfy the conditions for the validity (derivation) of *all* $AI_1$-$AI_6$ and $AI_8$. When the parameter is set to FALSE, the samples that satisfy the condition of the validity of a chosen CI are used, but they do not necessarily have to satisfy the conditions of *all* $AI_1$-$AI_6$ and $AI_8$.

Given these parameters, the function **makeTable** returns result matrices of the simulation (corresponding to Tables 3.2 to 3.8), where the elements are of the form of $\hat{r}_k\,(\widehat{CP}_{AI_k}, \widehat{CP}_{AI_1})$ defined via (3.3) and (3.1), for all $\alpha$, $n$ and $\gamma$ used, and a summary matrix of these result matrices, which consists of the ranges of $\widehat{CP}_{AI_k} - \widehat{CP}_{AI_1}$ (in %) and sample means of $\hat{r}_k$, $k = 2, ..., 8$, for each $\alpha$ used. Many parts of **makeTable** are borrowed from Tuzov (2014).

**nwise_summary** While the output of **makeTable** includes a summary matrix of the values of the ranges of $\widehat{CP}_{AI_k} - \widehat{CP}_{AI_1}$ (in %) and sample means of $\hat{r}_k$ of the results of the simulation for each $\alpha$ used, $k = 2, ..., 8$, the function **nwise_summary** returns a summary matrix of such values for each $n$ used. This function is used to construct Table 4.1.

```
################################################################################
#makeTable
################################################################################
makeTable<-function(CI,n,g,alpha,sigma=1,N=10000,seed=123,comparison=TRUE){
  if(CI==2) ci<-AI2
  else if(CI==3) ci<-AI3
  else if(CI==4) ci<-AI4
  else if(CI==5) ci<-AI5
  else if(CI==6) ci<-AI6
  else if(CI==7) ci<-AI7
  else if(CI==8) ci<-AI8
  else return("Choose a Valid number for CI")

  n.length<-length(n)
  g.length<-length(g)
  alpha.length<-length(alpha)
  ng.length<-n.length*g.length

  n.vec<-rep(n,times=g.length)
  g.vec<-rep(g,each=n.length)
  args<-matrix(c(n.vec,g.vec),ng.length,2)
  set.seed(seed)
  result_list<-list()
  summary_mat<-matrix(0,nrow=2,ncol=alpha.length)

  for(i in 1:alpha.length){
    ci.table<-function(args){
      n.cur<-args[1]
      g.cur<-args[2]
      alpha.cur<-alpha[i]
      sigma<-1
      if(comparison) add.cond<-TRUE
      else add.cond<-FALSE
      sample<-replicate(N,ci(n.cur,g.cur,alpha.cur,sigma,add.cond))
      return(paste(formatC(mean(sample[1,])/mean(sample[3,]),format="f",
          digits=3),"(",formatC(sum(sample[2,])/N,format="f",digits=3),",",
          formatC(sum(sample[4,])/N,format="f",digits=3),")",sep=""))
    }
    res.mat=matrix(apply(args,1,ci.table),nrow=n.length,ncol=g.length,F)
    res.list<-as.list(res.mat)
    num_res<-sapply(res.list,splitit,expr=("\\(|,|\\)"))
    mean_r<-mean(num_res[1,])
    diff_row<-num_res[2,]-num_res[3,]
    num_res<-rbind(num_res,diff_row)
    range1<-range(num_res[4,])
    range_perc<-range1*100
    summary_mat[1,i]<-paste(range_perc[1],"% - ",range_perc[2],"%")
    summary_mat[2,i]<-round(mean_r,3)
    results=cbind(n,res.mat)
    colnames(results)=c("n",g)
    result_list[[i]]<-results
    rm(results,res.mat,ci.table,res.list,num_res,mean_r,diff_row,range1,range
        _perc)
```

```
  }
  names(result_list)<-alpha
  colnames(summary_mat)<-alpha
  rownames(summary_mat)<-c("range of CP","r hat")
  result_list[["summary"]]<-summary_mat
  return(result_list)

}
################################################################################
#nwise_summary
################################################################################
nwise_summary<-function(res.list){
  res.length<-length(res.list)
  n<-nrow(res.list[[1]])
  nwise.list<-list()

  for(j in 1:n){
    nwise.list[[j]]<-numeric(0)
  }
  for(i in 1:(res.length-1)){
    awise.list<-as.list(res.list[[i]])
    nnames<-awise.list[(1:n)]
    awise.list<-awise.list[-(1:n)]
    num_res<-sapply(awise.list,splitit,expr=("\\(|,|\\)"))
    name.vec<-vector()
    for(j in 1:n){
      name.vec<-cbind(name.vec,nnames[[j]])
      nwise.list[[j]]<-cbind(nwise.list[[j]],num_res[,seq(j,ncol(num_res),4)
          ])
    }
  }
  nwise.matrix<-sapply(nwise.list,extract)
  rownames(nwise.matrix)<-c("range of CP","r hat")
  colnames(nwise.matrix)=name.vec
  return(nwise.matrix)
}
################################################################################
#Table making
#Before running this code, codes from Miscellaneous should be run first.
################################################################################
resultAI2<-makeTable(CI=2,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI2)
resultAI3<-makeTable(CI=3,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI3)
resultAI4<-makeTable(CI=4,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI4)
resultAI5<-makeTable(CI=5,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI5)
resultAI6<-makeTable(CI=6,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI6)
resultAI7<-makeTable(CI=7,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
```

```
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI7)
resultAI8<-makeTable(CI=8,n=c(50,100,300,1000),g=c(0.1,0.05,0.02),alpha=c
    (0.5,1.5,2,3,5),N=10000,seed=123,comparison=T)
nwise_summary(resultAI8[-(1:3)])
```

**Tables for the GME based CI's** Due to a long computational time for constructing the GME of (3.60) with a big $n$, instead of running one code as **makeTable**, we resort to building Tables 3.9 and 3.10 entry-by-entry, as it would take about 50 days for each of these tables if they were constructed at one go. Below is a sample code for one entry of Table 3.9 ($n = 100$, $\gamma = 0.1$, $\alpha = 1.5$, $\sigma = 1$). Before we simulate this entry, we generate the same number of samples that would have been generated up to this entry if the entire Table 3.9 was made with one execution of **makeTable**. Other entries of Table 3.9, as well as Table 3.10, are built in a similar fashion. We note in passing that even though $AI_9$ of (3.62) has a special condition for it to be valid, samples which violate such a condition did not arise in our simulations. $AI_{10}$ of (3.63) does not have a condition. The conditions of other CI's ($AI_2$-$AI_6$ and $AI_8$) are not applied on making Tables 3.9 and 3.10.

```
##############################################################################
#Tables for GME based CIs
##############################################################################
set.seed(123)
N<-500 #For the tables for AI9 and AI10, the number of repetition is 500.

dummy<-replicate(N,runif(50))  #Entry for alpha=0.5,n=50,g=0.1,sigma=1
dummy<-replicate(N,runif(100)) #Entry for alpha=0.5,n=100,g=0.1,sigma=1
dummy<-replicate(N,runif(200)) #Entry for alpha=0.5,n=200,g=0.1,sigma=1
dummy<-replicate(N,runif(300)) #Entry for alpha=0.5,n=300,g=0.1,sigma=1

dummy<-replicate(N,runif(50))  #Entry for alpha=0.5,n=50,g=0.0.05,sigma=1
dummy<-replicate(N,runif(100)) #Entry for alpha=0.5,n=100,g=0.0.05,sigma=1
dummy<-replicate(N,runif(200)) #Entry for alpha=0.5,n=200,g=0.0.05,sigma=1
dummy<-replicate(N,runif(300)) #Entry for alpha=0.5,n=300,g=0.0.05,sigma=1

dummy<-replicate(N,runif(50))  #Entry for alpha=0.5,n=50,g=0.02,sigma=1
dummy<-replicate(N,runif(100)) #Entry for alpha=0.5,n=100,g=0.02,sigma=1
dummy<-replicate(N,runif(200)) #Entry for alpha=0.5,n=200,g=0.02,sigma=1
dummy<-replicate(N,runif(300)) #Entry for alpha=0.5,n=300,g=0.02,sigma=1

dummy<-replicate(N,runif(50))  #Entry for alpha=0.5,n=50,g=0.1,sigma=1
```

```
dummy<-replicate(N,runif(100)) #Entry for alpha=0.5,n=100,g=0.1,sigma=1
dummy<-replicate(N,runif(200)) #Entry for alpha=0.5,n=200,g=0.1,sigma=1
dummy<-replicate(N,runif(300)) #Entry for alpha=1.5,n=300,g=0.1,sigma=1

dummy<-replicate(N,runif(50))  #Entry for alpha=1.5,n=50,g=0.05,sigma=1

result<-replicate(N,AI9(100,0.05,1.5,1)) #Our entry of interest
paste(formatC(mean(result[1,]),format="f",digits=8),",",formatC(mean(result
    [3,]),format="f",digits=8),",","(",formatC(sum(result[2,])/N,format="f",
    digits=8),",",formatC(sum(result[4,])/N,format="f",digits=8),")",sep="")
```

## A.5 Miscellaneous

The functions listed here have to be run in advance to use the codes given in Appendix.

**a.mme** Given a random sample $x$ from Pareto(I)($\alpha$,$\sigma$) and a value of $\sigma$, **a.mme** returns the MME of (3.52).

**a.mle** Given a random sample $x$ from Pareto(I)($\alpha$,$\sigma$) and a value of $\sigma$, **a.mle** returns the MLE of (2.2).

**rpareto** Given $n$, $\alpha$ and $\sigma$, **rpareto** returns $n$ number of random samples from Pareto(I)($\alpha$,$\sigma$).

**splitit** Given a string and an expression, the function **splitit** splits the string by the given expression and returns only the numbers in the string. For our purpose, **splitit** is used within the function **makeTable** and takes elements of the result matrices of the simulation as a parameter, in which each element is a string of the form of $\hat{r}_k\,(\widehat{CP}_{AI_k},\widehat{CP}_{AI_1})$, and returns the numbers $\hat{r}_k$, $\widehat{CP}_{AI_k}$ and $\widehat{CP}_{AI_1}$, without **(**, **)** or **,** in the elements.

**extract** The function **extract** is called within **nwise_summary**. It takes one result matrix from the output of the result matrices of the function **tableMake** as a parameter and returns the summary of the range of $\widehat{CP}_{AI_k} - \widehat{CP}_{AI_1}$ and the sample mean of $\hat{r}_k$ for each $n$ used in the matrix.

```
###############################################################################
#a.mme
###############################################################################
a.mme<-function(x,sigma){
  mme<-mean(x)/(mean(x)-sigma)
  return(mme)
}
###############################################################################
#a.mle
###############################################################################
a.mle<-function(x,sigma){
  n<-length(x)
  mle<-n/(sum(log(x/sigma)))
  return(mle)
}
###############################################################################
#rpareto
###############################################################################
rpareto<-function(n,a,b){
  #inverse Transformation Method
  if(a<0 || b<0){
    return("a and b should be positive")
  }
  inv.cdf<-function(u){
    x<-b/((1-u)^(1/a))
  }
  U<-runif(n,0,1)
  X<-inv.cdf(U)
  return(X)
}
###############################################################################
#splitit
###############################################################################
splitit<-function(string,expr){
  list_res<-strsplit(string,expr)
  vector_res<-as.vector(list_res[[1]])
  numeric_res<-as.numeric(vector_res)
  return(numeric_res)
}
###############################################################################
#extract
###############################################################################
extract<-function(x){
  mean_r<-mean(x[1,])
  diff_row<-x[2,]-x[3,]
  range<-range(diff_row)
  range_perc<-range*100
  range_str<-paste(range_perc[1],"% - ",range_perc[2],"%")
  mat<-matrix(c(mean_r,range_str),2,1)
  return(mat)
}
```

# Bibliography

Arnold, B. C. (1983). *Pareto Distributions*. International Co-operative Publishing House. (Cited on pages 1, 2, 8, 11, 12, 13 and 39.)

Brazauskas, V. and R. Serfling (2000a). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal 4*(4), 12–27. (Cited on pages 2, 3, 4, 5, 45, 46, 47, 49, 59 and 61.)

Brazauskas, V. and R. Serfling (2000b). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes 3*(3), 231–249. (Cited on pages 54 and 55.)

Brazauskas, V. and R. Serfling (2003, November). Favorable estimators for fitting Pareto models: A study using goodness-of-fit measures with actual data. *ASTIN Bulletin 33*(2), 365–381. (Cited on page 2.)

Casella, G. and R. L. Berger (2002). *Statistical Inference*. Brooks/Cole Cengage Learning. (Cited on page 9.)

Csörgő, M., B. Szyszkowicz, and Q. Wang (2003). Donsker's theorem for self-normalized partial sums processes. *The Annals of Probability 31(3)*, 1228–1240. (Cited on page 24.)

Csörgő, S. and L. Horváth (1985). The baboons come down from the trees quite normally. *Mathematical statistics and applications B*, 95–106. (Cited on pages 28 and 32.)

Giné, E., F. Götze, and D. M. Mason (1997). When is the Student $t$-statistic asymptotically standard normal? *The Annals of Probability 25(3)*, 1514–1531. (Cited on page 19.)

Gut, A. (2009). *An Intermediate Course in Probability.* Springer New York. (Cited on page 17.)

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous univariate distributions* (2nd ed.). New York Wiley & Sons. (Cited on page 1.)

Lévy, P. (1937). *Théorie de l'addition des variables aléatoires.* Gauthier-Villars. (Cited on page 18.)

Martsynyuk, Y. V. (2013). On the generalized domain of attraction of the multivariate normal law and asymptotic normality of the multivariate Student $t$-statistic. *Journal of Multivariate Analysis 114*, 402–411. (Cited on page 20.)

Martsynyuk, Y. V. and E. Tuzov (2016). Exploring functional CLT confidence intervals for a population mean in the domain of attraction of the normal law. *Acta Mathematica Hungarica 148*(2), 493–508. (Cited on pages 3, 24 and 25.)

Pinsky, M. and S. Karlin (2011). *An Introduction to Stochastic Modelling*. Academic Press. (Cited on page 32.)

Quandt, R. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika 10(1)*, 55–82. (Cited on page 53.)

Tuzov, E. (2014). Exploring functional asymptotic confidence intervals for a population mean. Master's thesis, University of Manitoba, Winnipeg, Manitoba, Canada. (Cited on pages 66 and 73.)