Asymptotic Properties of the Hill Estimator

Jaakko Pere

School of Science

Bachelor's thesis
Espoo 23.8.2018

Supervisor

Ph.D Pauliina Ilmonen

Advisor

M.Sc Matias Heikkilä

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Jaakko Pere

Title Asymptotic Properties of the Hill Estimator

Degree programme Technical Physics and Mathematics

Major Mathematics and Systems Analysis **Code of major** SCI3025

Supervisor Ph.D Pauliina Ilmonen

Advisor M.Sc Matias Heikkilä

Date 23.8.2018 **Number of pages** 22+2 **Language** English

Abstract

Your abstract in English. Keep the abstract short. The abstract explains your research topic, the methods you have used, and the results you obtained.

The abstract text of this thesis is written on the readable abstract page as well as into the pdf file's metadata via the \thesisabstract macro (see above). Write here the text that goes onto the readable abstract page. You can have special characters, linebreaks, and paragraphs here. Otherwise, this abstract text must be identical to the metadata abstract text.

If your abstract does not contain special characters and it does not require paragraphs, you may take advantage of the abstracttext macro (see the comment below).

Keywords For keywords choose, concepts that are, central to your, thesis

Preface

I want to thank Professor Pirjo Professori and my instructor Dr Alan Advisor for their good and poor guidance.

Otaniemi, 24.4.2018

Eddie E. A. Engineer

Contents

Symbols and abbreviations

Symbols

$x^* = \sup\{x : F(x) < 1\}$	right endpoint of the distribution
γ	extreme value index
$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$	left-continuous inverse
U	left-continuous inverse of $\frac{1}{1-F}$
$\mathbb{1}(p) = \begin{cases} 1, \text{if p is true} \\ 0, \text{otherwise} \end{cases}$	indicator fuction
$X_{i,n}$	ith order statistic
λ	Lebesque measure
$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$	limit supremum of a sequence of sets A_n
$f \in RV_\alpha$	f is an regularly varying function with index α
G	extreme value distribution
$f \in D(G_\gamma)$	f is in the maximum domain of attraction of G

Abbreviations

cdf	cumulative distribution function
i.d.d.	independent and identically distributed
a.s.	almost surely

1 Introduction

2 Backround

2.1 Fisher-Tippett-Gnedenko Theorem and Domains of Attraction

First approach to study the behavior of extreme events could be to find limiting distribution of the sample maxima $M_n = \max(X_1, X_2, ..., X_n)$. Here $X_1, X_2, ..., X_n$ are i.d.d. random variables from cdf F_X. Function for the cdf of M_n can be easily derived, since $X_1, X_2, ..., X_3$ are i.d.d.

$$P(\max(X_1, X_2, ..., X_n) \leq x) = P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) =$$
$$P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x) = F^n(x).$$

Now it can be shown that this approach is not very useful since

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 0, x < x^* \\ 1, x \geq x^*. \end{cases}$$

To achieve a nondegerate distribution it is necessary to normalize the sample maxima M_n. After normalization a nondegenate distribution is gained as stated in the Fisher-Tippett-Gnedenko Theorem [2], [3].

Theorem 2.1. *There exists real constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(ax + b), \tag{1}$$

where

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}), \gamma \neq 0 \\ \exp(-e^{-x}), \gamma = 0, \end{cases}$$

for all x with $1 + \gamma x > 0$ where $\gamma \in \mathbb{R}$.

If F fullfills the equation 1 for some $\gamma \in \mathbb{R}$ then it is said that F is in the maximum domain of attraction of G_γ i.e. $F \in D(G_\gamma)$. Considering the Hill estimator we are especially interested in the case $F \in D(G_{\gamma>0})$. It turns out that $F \in D(G_{\gamma>0})$ is equivalent to the fact that function $1 - F$ is regularly varying with index $-\frac{1}{\gamma}$. [4]

Theorem 2.2. *Cdf F is in the maximum domain of attraction of the extreme value distribution G_γ with $\gamma > 0$ if and only if $x^* = \infty$ and*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\frac{1}{\gamma}}, x > 0. \tag{2}$$

In addition, condition 2 can be written in different form with the U function [4].

Corollary 2.3. *Condition 2 is equivalent to*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, x > 0. \tag{3}$$

Above equation implies that U is regularly varying with index γ if $F \in D(G_{\gamma>0})$.

2.2 Regularly Varying Functions

In section 2.1 we saw that if $F \in D(G_{\gamma>0})$ then U is regularly varying function. Regularly varying functions have some useful properties that are needed to prove the consistency of the Hill estimator. Let's define regularly varying functions properly [4]:

Definition 2.4. *A Lebesque measurable function $f : \mathbb{R}^+ \to \mathbb{R}$ that is eventually positive is regularly varying if for some index $\alpha \in \mathbb{R}$,*

$$\lim_{x\to\infty} \frac{f(tx)}{f(t)} = x^{\alpha}, \quad x > 0. \tag{4}$$

If function f is regularly varying with index $\alpha = 0$ then f is called slowly varying. For a slowly varying function the limit relation 4 can be written in different form with function $F = \log f(e^x)$:

$$\lim_{t\to\infty} F(t+x) - F(x) = 0. \tag{5}$$

The above argument is true, since

$$F(t+x) - F(t) = \log f(e^{t+x}) - \log f(e^t) = \log\left(\underbrace{\frac{f(e^t e^x)}{f(e^t)}}_{\to 1}\right) \to 0$$

as $t \to \infty$. The alternative form for slow variation 5 is used in the proof of the uniform convergence.

Theorem 2.5. *If $f \in RV_\alpha$ then the convergence in the equation 4 is uniform .*

$$\lim_{t\to\infty} \sup_{x\in[a,b]} \left|\frac{f(tx)}{f(t)} - x^{\alpha}\right| = 0,$$

for $0 < a < b < \infty$.

Proof. For the proof it can be assumed that $\alpha = 0$. If this isn't the case replace $f(x)$ by $f(x)x^{-\alpha}$. Suppose there exists sequences $t_n \to \infty$, $x_n \to 0$ as $n \to \infty$ such that

$$\left|\frac{f(t_n x_n)}{f(t_n)} - 1\right| > \delta$$

for all $n \in \mathbb{N}$ and some $\delta > 0$. An equivalent condition can be formulated with function $F(x) = \log f(e^x)$(see equation 5):

$$|F(t_n + x_n) - F(t_n)| > \delta \tag{6}$$

with possibly different x_n, t_n and δ. Let's define sets

$$\begin{aligned} Y_{1,n} &= \left\{ y \in J : |F(t_n + y) - F(t_n)| > \frac{\delta}{2} \right\}, \\ Y_{2,n} &= \left\{ y \in J : |F(t_n + x_n) - F(t_n + y)| > \frac{\delta}{2} \right\} \quad \text{and} \\ Z_n &= \left\{ z : |F(t_n + x_n) - F(t_n + x_n - z)| > \frac{\delta}{2}, x_n - z \in J \right\} \\ &= \{z : x_n - z \in Y_{2,n}\} \end{aligned}$$

where $J \subset \mathbb{R}$ is a finite interval. Next we will prove that if the equation 6 holds then pointwise convergence $\lim_{t\to\infty} F(t + x_0) - F(t) = 0$ cannot hold. Pointwise convergence does not hold if some x_0 is included in infinitely many $Y_{1,n}$. Reason for this is that

$$n \geq n_\varepsilon \Rightarrow |F(t + x_0) - F(t)| < \varepsilon, \forall \varepsilon > 0, \exists n_\varepsilon \in \mathbb{N} \tag{7}$$

cannot hold if x_0 is included in infinitely many $Y_{1,n}$. This can be noticed by comparing equation 7 and the condition of $Y_{1,n}$. Similarly if x_0 is included in infinitely many Z_n then pointwise convergence cannot hold, since the condition in Z_n can be written as

$$\begin{aligned} &\left| F(\underbrace{t_n + x_n}_{=u_n}) - F(\underbrace{t_n + x_n}_{=u_n} \overbrace{-z}^{=x_0}) \right| > \frac{\delta}{2} \\ &\Leftrightarrow |F(u_n + x_0) - F(u_n)| > \frac{\delta}{2} \end{aligned}$$

where $u_n \to \infty$.

Notice that $Y_{1,n} \cup Y_{2,n} = J$, since by the equation 6 and triangle inequality we have

$$\begin{aligned} \delta < |F(t_n + x_n) - F(t_n)| &= |(F(t_n + x_n) - F(t_n + y)) + (F(t_n + y) - F(t_n))| \\ &\leq |(F(t_n + x_n) - F(t_n + y))| + |(F(t_n + y) - F(t_n))| \\ \Rightarrow |(F(t_n + x_n) - F(t_n + y))| > \frac{\delta}{2} &\vee |(F(t_n + y) - F(t_n))| > \frac{\delta}{2}. \end{aligned}$$

Additionally $Y_{1,n}$, $Y_{2,n}$ and J are measurable sets. So by subadditivity of the Lebesque measure we have $\lambda(Y_{1,n}) \geq \frac{\lambda(J)}{2} \vee \lambda(Y_{2,n}) \geq \frac{\lambda(J)}{2}$. By the translation property of the Lebesque measure $\lambda(Z_n) = \lambda(Y_{2,n})$ holds. Thus $\lambda(Y_{1,n}) \geq \frac{\lambda(J)}{2} \vee \lambda(Z_n) \geq \frac{\lambda(J)}{2}$ infinitely often. All $Y_{1,n}$ are subsets of finite interval since $Y_{1,n} \subset J$ for all n. Similarly all Z_n are subset of a finite interval since $x_n \to 0$. Hence by Fatou's lemma [1]:

$$\begin{aligned} \lambda(\limsup Y_{1,n}) \geq \limsup \lambda(Y_{1,n}) \geq \frac{\lambda(J)}{2} \quad &\vee \\ \lambda(\limsup Z_n) \geq \limsup \lambda(Z_n) \geq \frac{\lambda(J)}{2}&. \end{aligned}$$

Since at least one of the measures $\lambda(\limsup Y_{1,n})$ or $\lambda(\limsup Z_n)$ is greater than zero, we have some x_0 that is contained in infinitely many $Y_{1,n}$ or Z_n. This was the desired contradiction. $\square$

With uniform convergence it can be proved that all the regularly varying functions are in certain form:

Theorem 2.6 (Karamata's representation theorem)**.** *If $f \in RV_\alpha$ then there exists measurable functions $a : \mathbb{R} \to \mathbb{R}^+$ and $c : \mathbb{R} \to \mathbb{R}^+$ with*

$$\lim_{t\to\infty} c(t) = c_0 \text{ and } \lim_{t\to\infty} a(t) = \alpha$$

and $t_0 \in \mathbb{R}^+$ such that for $t > t_0$

$$f(t) = c(t) \exp\left(\int_{t_0}^{t} \frac{a(s)}{s} ds\right) \tag{8}$$

Conversely, if 2.6 holds, then $f \in RV_\alpha$.

For the proof of the above theorem following lemma is needed.

Lemma 2.7. *Suppose $f \in RV_\alpha$. There exists $t_0 > 0$ such that $f(t)$ is positive and locally bounded for $t \geq t_0$. If $\alpha \geq -1$ then*

$$\lim_{t\to\infty} \frac{tf(t)}{\int_{t_0}^{t} f(s)ds} = \alpha + 1. \tag{9}$$

If $\alpha < -1$ or $\alpha = -1$ and $\int_0^\infty f(s)ds < \infty$, then

$$\lim_{t\to\infty} \frac{tf(t)}{\int_{t}^{\infty} f(s)ds} = -\alpha - 1. \tag{10}$$

Conversely, if 9 holds for $-1 \leq \alpha < \infty$ or 10 holds for $-\infty < \alpha < -1$, then $f \in RV_\alpha$.

Next we prove the above lemma.

Proof. First we prove the equation 9. Suppose that $f \in RV_\alpha$. Then by theorem 2.5 there exists t_0 and c such that $f(tx)/t < c$ when $t \geq t_0$, $x \in [1, 2]$. Then for $t \in [2^n t_0, 2^{n+1} t_0]$ we have

$$\frac{f(t)}{f(t_0)} = \frac{f(t)}{f(2^{-1}t)} \frac{f(2^{-1}t)}{f(2^{-2}t)} ... \frac{f(2^{-n}t)}{f(t_0)} < c^{n+1}. \tag{11}$$

Equation 11 is true since every fraction can be written as $f(tx)/f(t)$. This implies that for $t \geq t_0$ $f(t)$ is both locally bounded and $\int_{t_0}^{t} f(s)ds < \infty$. Consider a function $F(t) = \int_{t_0}^{t} f(s)ds$.We start by proving that $\lim_{t\to\infty} F(t) = \infty$ when $\alpha > -1$. First notice that $f(2s) \geq 2^{-1}f(s)$ for sufficiently large s. For $n \geq n_0$

$$\int_{2^n}^{2^{n+1}} f(s)ds = 2\int_{2^{n-1}}^{2^n} f(2s)ds \geq \int_{2^{n-1}}^{2^n} f(s)ds \tag{12}$$

by the change on variables. Then by induction we have

$$\int_{2^n}^{2^{n+1}} f(s)ds \geq \int_{2^{n_0}}^{2^{n_0+1}} f(s)ds = C > 0. \tag{13}$$

Thus

$$\int_{2^{n_0}}^{\infty} f(s)ds = \sum_{n=n_0}^{\infty} \int_{2^n}^{2^{n+1}} f(s)ds \geq \sum_{n=n_0}^{\infty} \int_{2_0^n}^{2^{n_0+1}} f(s)ds = \sum_{n=n_0}^{\infty} C = \infty \tag{14}$$

Next we prove that $F \in RV_{\alpha+1}$ for $\alpha > -1$. Let $\varepsilon > 0$ and $t_1 = t_1(\varepsilon)$. Then $f(xt) < (1+\varepsilon)x^\alpha f(t)$ for $t > t_1$. Since $\lim_{t\to\infty} F(t) = \infty$,

$$\frac{F(tx)}{F(t)} = \frac{\int_{t_0}^{tx} f(s)ds}{\int_{t_0}^{t} f(t)ds} \sim \frac{\int_{t_1 x}^{tx} f(s)ds}{\int_{t_1}^{t} f(t)ds} = \frac{x\int_{t_1}^{t} f(xs)ds}{\int_{t_1}^{t} f(t)ds} < \frac{x\int_{t_1}^{t} (1+\varepsilon)x^\alpha f(s)ds}{\int_{t_1}^{t} f(t)ds} = (1+\varepsilon)x^{\alpha+1}$$

by the change of variables. A similar lower bound for $F(tx)/F(t)$ can be derived by using $f(xt) < (1-\varepsilon)x^\alpha f(t)$ as $t > t_1$. So we have that $F \in RV_{\alpha+1}$ for $\alpha > -1$. In the case $\alpha = -1$ and $F(t) \to \infty$ same proof applies. If $\alpha = -1$ and $F(t)$ has a finite limit and $F \in RV_0$. Now for all α

$$\begin{aligned}\frac{F(xt)-F(t)}{tf(t)} &= \frac{1}{tf(t)}\int_t^{tx} f(u)du = \frac{t}{tf(t)}\int_1^x f(ut)du = \int_1^x \frac{f(ut)}{f(t)}du \\ &\to \int_1^x u^\alpha du = \frac{x^{\alpha+1}-1}{\alpha+1}, \quad t\to\infty\end{aligned}$$

by the theorem 2.5 and change of variables. On the other hand

$$\begin{aligned}\frac{F(xt)-F(t)}{tf(t)} = \frac{F(t)}{tf(t)}\left(\underbrace{\frac{F(tx)}{F(t)}}_{\to x^{\alpha+1}} -1\right) &\to \frac{x^{\alpha+1}-1}{\alpha+1} \\ \Rightarrow \lim_{t\to\infty}\frac{tf(t)}{F(t)} &= \alpha+1\end{aligned}$$

Now we have proven 9. Next we prove equation 10. Let's define

$$G(t) = \int_t^\infty f(s)ds$$

In the case $\alpha < -1$ there exists $\delta > 0$ such that $f(2s) \leq 2^{-1-\delta}f(s)$ for sufficiently large s. Now we can prove the finiteness of $\lim_{t\to\infty} G(t)$ in a similar way as the infinitess of $\lim_{t\to\infty} F(t)$ in equations 12, 13 and 14. For sufficiently large n_1

$$\begin{aligned}\int_{2^n}^{2^{n+1}} f(s)ds = 2\int_{2^{n-1}}^{2^n} f(s)ds \leq 2^{-\delta}\int_{2^{n-1}}^{2^n} f(s)ds \leq \\ \ldots \leq 2^{-\delta(n-n_1)}\int_{2^{n_1}}^{2^{n_1+1}} f(s)ds = 2^{-\delta(n-n_1)}C'\end{aligned}$$

by induction and change of variables. Then

$$\begin{aligned}\int_{2^{n_1}}^{\infty} f(s)ds = \sum_{n=n_1}^{\infty} \int_{2^n}^{2^{n+1}} f(s)ds &\le C' \sum_{n=n_1}^{\infty} 2^{-\delta(n-n_1)} \\ &= C' \sum_{k=0}^{\infty} \left(\frac{1}{2^\delta}\right)^k = \frac{C'}{1-1/2^\delta} < \infty,\end{aligned}$$

Now rest of the proof is analogous. Next we prove the converse results. Suppose that equation 9 holds. Let's define a function

$$b(t) = t\frac{f(t)}{F(t)}$$

Without loss of generality we may suppose that $f(t) > 0$ and $t > 0$. Integrating both sides of $b(t)/t = f(t)/F(t)$ we obtain for some real c_1 and for all $x > 0$

$$\int_1^x \frac{b(t)}{t} dt = \log F(x) + c_1, \tag{15}$$

since by change of variables

$$\int_1^x \frac{f(t)}{F(t)} dt = \int_{F(1)}^{F(x)} \frac{1}{u} du = \log F(x) + \underbrace{\log F(1)}_{=c_1}.$$

From the equation 15 we have

$$F(t) = \exp\left(\int_1^x \frac{b(t)}{t} dt - c_1\right) = \underbrace{\exp(-c_1)}_{=c} \exp\left(\int_1^x \frac{b(t)}{t} dt\right) = c\exp\left(\int_1^x \frac{b(t)}{t} dt\right).$$

Then by using the definition of f again

$$\begin{aligned} f(x) = x^{-1}b(x)F(x) = cb(x)\exp\left(-\int_1^x \frac{1}{t}\right)\exp\left(\int_1^x \frac{b(t)}{t}\right) \\ = cb(x)\exp\left(\int_1^x \frac{b(t)-1}{t} dt\right), \end{aligned} \tag{16}$$

for all $x > 0$. Hence for all $x, t > 0$

$$\begin{aligned}\frac{f(tx)}{f(t)} = \frac{b(tx)\exp\left(\int_1^{tx} \frac{b(s)-1}{s} ds\right)}{b(tx)\exp\left(\int_1^{t} \frac{b(s)-1}{s} ds\right)} &= \frac{b(tx)}{b(t)}\exp\left(\int_1^{tx} \frac{b(s)-1}{s} ds - \int_1^t \frac{b(s)-1}{s} ds\right) \\ = \frac{b(tx)}{b(t)}\exp\left(\int_t^{tx} \frac{b(s)-1}{s} ds\right) &= \frac{b(tx)}{b(t)}\exp\left(\int_1^{x} \frac{b(ts)-1}{s} ds\right),\end{aligned}$$

by the change of variables. By the assumption (equation 9) $b(t) \to \alpha + 1$ so $b(tx)/b(t) \to 1$. For sufficiently large t

$$\exp\left(\int_1^x \frac{b(ts)-1}{s} ds\right) \approx \exp\left(\int_1^x \frac{\alpha}{s} ds\right) = \exp(\alpha \log x) = x^\alpha$$

The last statement (equation 10 implies that $F \in RV_\alpha$) can be proved in a similar way. $\square$

Next we prove the theorem 2.6.

Proof. Suppose $f \in RV_\alpha$. The function $t^{-\alpha}f(t)$ is slowly varying and

$$t^{-\alpha}f(t) = cb(t)\exp\left(\int_1^t \frac{b(s)-1}{s}ds\right)$$

by the equation 16. Now by lemma 2.7 $b(t) \to 1$ and function $t^{-\alpha}f(t)$ has the representation as in theorem 2.6 with $a(t) = b(t) - 1$ and $c(t) = cb(t)$. Then

$$f(t) = c(t)t^\alpha \exp\left(\int_{t_0}^t \frac{a(s)}{s}ds\right)$$

Notice that we can write t^α as $\exp\left(\int_1^t \frac{\alpha}{s}ds\right)$. Then f has the form

$$\begin{aligned} f(t) &= c(t)\exp\left(\int_{t_0}^t \frac{a(s)}{s}ds + \int_1^t \frac{\alpha}{s}ds\right) \\ &= c(t)\exp\left(\int_{t_0}^t \frac{a(s)+\alpha}{s}ds + \int_1^{t_0} \frac{\alpha}{s}ds\right) \\ &= c(t)\exp\left(\int_{t_0}^t \frac{a(s)+\alpha}{s}ds\right)\exp\left(\log t_0^\alpha\right) \end{aligned}$$

$$= \underbrace{t_0^\alpha c(t)}_{=c'}\exp\left(\int_{t_0}^t \frac{\overbrace{a(s)+\alpha}^{=a'}}{s}ds\right) \tag{17}$$

From the equation 17 it can be seen that $f(t)$ has the same representation as in the theorem 2.6 when a is replaced by a' and c is replaced by c'.

□

Next corollary will be crucial in the proof of the consistency of the Hill estimator.

Corollary 2.8. *Suppose $f \in RV_\alpha$. If $\varepsilon, \delta > 0$ are arbitrary, there exists $t_0 = t_0(\varepsilon, \delta)$ such that for $t \geq t_0$, $tx \geq t_0$,*

$$(1-\varepsilon)x^{\alpha-\delta} < \frac{f(tx)}{f(t)} < (1+\varepsilon)x^{\alpha+\delta}$$

Above corollary follows from the theorem 2.6.

Proof. By the theorem 2.6

$$\frac{f(tx)}{f(t)} = \frac{c(tx)}{c(t)}\exp\left(\int_1^x \frac{a(st)}{s}ds\right)$$

The function c(t) converges to a constant. Hence $c \in RV_0$ so $c(tx)/c(t) \to 1$ as $t \to \infty$. Furthermore, $a(s) \to \alpha$ as $t \to \infty$. Now we can choose such a t_0 that $\alpha - \delta < a(st) < \alpha - \delta$ and $1 - \varepsilon < \frac{c(tx)}{c(t)} < 1 + \varepsilon$. This implies that

$$\begin{aligned}(1-\varepsilon)\int_1^x \frac{\alpha-\delta}{s}ds < \frac{f(tx)}{f(t)} < (1+\varepsilon)\int_1^x \frac{\alpha+\delta}{s}ds \\ \Rightarrow (1-\varepsilon)\exp\left(\log\left(x^{\alpha-\delta}\right)\right) < \frac{f(tx)}{f(t)} < (1+\varepsilon)\exp\left(\log\left(x^{\alpha+\delta}\right)\right) \\ \Rightarrow (1-\varepsilon)x^{\alpha-\delta} < \frac{f(tx)}{f(t)} < (1+\varepsilon)x^{\alpha+\delta}\end{aligned}$$

□

3 Hill Estimator

3.1 Consistency

The following theorem states that Hill estimator is consistent i.e. estimator converges in probability to extreme value index. [5]

Theorem 3.1. *Let $X_1, X_2, ...$ be i.d.d. variables with cdf F_X. Suppose $F_X \in D(G_\gamma)$ with $\gamma > 0$. Then as $n \to \infty$, $k = k(n) \to \infty$, $\frac{k}{n} \to 0$,*

$$\hat{\gamma}_H \xrightarrow{p} \gamma.$$

For the proof of the above theorem following lemmas are needed, firstly the Renyi's representation [7].

Lemma 3.2. *If $E_1, E_2, ...$ are i.d.d. random variables from the standard exponential distribution and $E_{1,n} \leq E_{2,n} \leq ... \leq E_{n,n}$ then for $k \leq n$ we have*

$$\Big(E_{1,n}, E_{2,n}, ..., E_{k,n}\Big) \stackrel{d}{=} \Big(\frac{E_1^*}{n}, \frac{E_1^*}{n} + \frac{E_2^*}{n-1}, ..., \frac{E_1^*}{n} + \frac{E_2^*}{n-1} + ... + \frac{E_k^*}{n-k+1}\Big),$$

where $E_1^*, E_2^*, ...$ are i.d.d. random variables from standard exponential distribution.

Secondly the lemma about the order statistics of Pareto distribution is necessary [4].

Lemma 3.3. *Let $Y_1, Y_2, ...$ be i.d.d. random variables from Pareto distribution $F_Y(y) = 1 - \frac{1}{y}$, $y \geq 0$ and let $Y_{1,n} \geq Y_{2,n} \geq ... \geq Y_{n,n}$ be the nth order statistics. Then with such $k = k(n)$ that $k \to \infty$, $\frac{k}{n} \to 0$ as $n \to \infty$,*

$$\lim_{n \to \infty} Y_{n-k,n} = \infty \quad a.s.$$

Last lemma we need says that U(Y) is equal in distribution to X, where Y is random variable from Pareto distribution and X is random variable from some distribution F_X.

Lemma 3.4. *Let Y be random variable from Pareto distribution $F_Y = 1 - \frac{1}{y}$, $y \geq 0$ Let X be random variable with cdf F_X then $U(Y) \stackrel{d}{=} X$.*

Next we prove the lemma 3.3. Proof of the lemma 3.2 is omitted here.

Proof. Let us assume that $Y_{n-k,n} < r$ for some $r > 0$ infinitely often. In other words

$$\frac{k}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Y_i > Y_{n-k,n}) > \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Y_i > r).$$

Now the left side of the equation converges to zero, since

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Y_i > Y_{n-k,n}) = \lim_{n\to\infty} \frac{k}{n} = 0.$$

But the right side converges to $1/r$ almost surely, since

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Y_i > r) \xrightarrow{a.s} P(Y_i > r) = 1 - F_Y(r) = \frac{1}{r}$$

by the strong law of large numbers [8]. So the assumption cannot hold which implies that

$$P(\lim_{n\to\infty} Y_{n-k,n} = \infty) = 1.$$

□

Now we prove the last lemma 3.4 that is needed for the proof of theorem 3.1

Proof. Let's study the condition $U(Y) \leq a, a \in \mathbb{R}$.

$$\begin{aligned} &U(Y) \leq a \\ \Leftrightarrow \inf &\Big\{x : \frac{1}{1 - F_X(x)} \geq Y\Big\} \leq a \\ \Leftrightarrow \inf &\Big\{x : 1 - \frac{1}{Y} \leq F_X(x)\Big\} \leq a \end{aligned} \tag{18}$$

Let $S = \Big\{x : 1 - \frac{1}{Y} \leq F_X(x)\Big\}$ and $b = \inf S$. Notice that F is increasing and right-continuous, since F is a cdf. So S is an interval of form $[b, \infty)$ or (b, ∞), since F is increasing. Let's define a sequence $x_n = b + \frac{1}{n}, n \in \mathbb{N}$. Notice that $x_n \to b$ and $x_n \in S$ for all n. Now right-continuity implies that $b \in S$ i.e S is an interval $[b, \infty)$. Additionally $a \in S$ since $a \geq b$ so a satisfies the condition $1 - \frac{1}{Y} \leq F(a)$. Therefore the equation 18 implies

$$U(Y) \leq a \Leftrightarrow 1 - \frac{1}{Y} \leq F_X(a) \Leftrightarrow Y \leq \frac{1}{1 - F(a)},$$

So now from the cdf of U(Y) we have

$$\begin{aligned} F_{U(Y)} = P(U(Y) \leq x) = P\Big(Y \leq \frac{1}{1 - F_X(x)}\Big) &= F_Y\Big(\frac{1}{1 - F_X(x)}\Big) \\ &= 1 - \Big(\frac{1}{1 - F_X(x)}\Big)^{-1} = F_X(x). \end{aligned}$$

□

Now we are equipped to prove the theorem 3.1.

Proof. $F \in D(G_{\gamma>0})$ is equivalent to the fact that $U \in RV_\gamma$ i.e.

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma.$$

From the uniform convergence of the regularly varying functions follows that for $x > 1$ and $t \geq t_0$,

$$(1-\varepsilon)x^{\gamma-\delta} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\gamma+\delta},$$

for all $\varepsilon > 0$ and $\delta > 0$. By taking natural logarithm from both sides of the equation above, it can be written as

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\log(x) &< \log(U(tx)) - \log(U(t)) \\ &< \log(1+\varepsilon) + (\gamma+\delta)\log(x). \end{aligned} \tag{19}$$

If $Y_1, Y_2, ...$ are i.d.d random variables from Pareto distribution with cdf $F_Y(y) = 1 - \frac{1}{y}$ then $U(Y_i) \stackrel{d}{=} X_i$ as stated in lemma 3.4. Hence it is sufficient to prove the result for $\hat{\gamma}_H = \frac{1}{k}\sum_{i=0}^{k-1} \log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n}))$. For $t = Y_{n-k,n}$ and $x = \frac{Y_{n-i,n}}{Y_{n-k,n}}$ equation 19 has the form

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) &< \log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n})) \\ &< \log(1+\varepsilon) + (\gamma+\delta)\log(\frac{Y_{n-i,n}}{Y_{n-k,n}}). \end{aligned} \tag{20}$$

Notice that we can replace t with $Y_{n-k,n}$ because we can always find some n_0 such that $Y_{n_0-k,n_0} \geq t_0$ according to lemma 3.3. Furthermore, $Y_{n-i,n}$ is greater than $Y_{n-k,n}$ always when $i < k$. Therefore x can be replaced with $\frac{Y_{n-i,n}}{Y_{n-k,n}}$.

Equation 20 applies for every $i = 0, 1, 2, ..., k-1$. Thus we can write

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) &< \frac{1}{k}\sum_{i=0}^{k-1}\log(U(Y_{n-i,n})) - \log(U(Y_{n-k,n})) \\ &< \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right). \end{aligned}$$

The term in the middle is the hill estimator $\hat{\gamma}_H$, hence above becomes

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) &< \hat{\gamma}_H \\ &< \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right). \end{aligned}$$

Now it is sufficient to only prove that

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)\xrightarrow{p}1.$$

$\log(Y_i)$ has a standard exponential distribution, since

$$F_{\log(Y_i)}(x)=P(\log(Y_i)<x)=P(e^{\log(Y_i)}<e^x)=P(Y_i<e^x)=F_Y(e^x)=1-e^{-x}.$$

Therefore we can write

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)=\frac{1}{k}\sum_{i=0}^{k-1}E_{n-i,n}-E_{n-k,n},$$

where $E_1, E_2, ...$ are i.d.d. random variables from standard exponential distribution. Now Renyi's representation 3.2 implies

$$\begin{aligned}
&\left\{E_{n-i,n}-E_{n-k,n}\right\}_{i=0}^{k-1}\\
&\overset{d}{=}\Bigg\{\left(\frac{E_1^*}{n}+\frac{E_2^*}{n-1}+...+\frac{E_{n-(i+1)}^*}{n-(n-(i+1))+1}+\frac{E_{n-i}^*}{n-(n-i)+1}\right)\\
&\qquad-\left(\frac{E_1^*}{n}+\frac{E_2^*}{n-1}+...+\frac{E_{n-k}^*}{n-(n-k)+1}\right)\Bigg\}_{i=0}^{k-1}\\
&=\left\{\frac{E_{n-i}^*}{i+1}+\frac{E_{n-(i+1)}^*}{i+2}+...+\frac{E_{n-(k-2)}^*}{k-1}+\frac{E_{n-(k-1)}^*}{k}\right\}_{i=0}^{k-1}\\
&\overset{d}{=}\left\{E_{k-i,k}\right\}_{i=0}^{k-1}.
\end{aligned}$$

Consequently we have

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)\overset{d}{=}\frac{1}{k}\sum_{i=0}^{k-1}E_{k-i,k}=\frac{1}{k}\sum_{i=0}^{k-1}E_i\xrightarrow{p}E[E_i]=1$$

by the weak law of large numbers [8]. Notice that the expected value of a standard exponential is one.

□

To note there is also converse version of the theorem 3.1 [6]. Proof is omitted here.

Theorem 3.5. *Let $X_1, X_2, ...$ be i.i.d. variables from cdf F. Suppose that for some sequence of integers $k=k(n)\to\infty$, $k(n)/n\to 0$ and $k(n+1)/k(n)\to 1$ as $n\to\infty$*

$$\hat{\gamma}_H\to\gamma>0.$$

Then $F\in D(G_\gamma)$

4 Simulations

The consistency of the Hill estimator is tested with simulations. We simulate from Pareto and Cauchy distributions. Cdf of the Pareto distribution is $F(x) = 1 - \left(\frac{1}{x}\right)^3$ and the cdf of the Cauchy distribution is $F(x) = \frac{1}{\pi}\arctan(x)$. Corresponding extreme value indexes are 1/3 for Pareto distribution and 1 for Cauchy distribution. This can be checked with equation 2 or 3.We simulate N sets of n_i observations from both distributions where $n_i = 100 + 50(i-1), i = 1, 2, ..., 399$. Hence the sample size is a vector from 100 to 20000 with steps of 50. We mark this vector with $\mathbf{n} = (100, 150, ..., 20000)$. The number of samples N is set to 2000. So for each n_i we calculate N number of estimates $\hat{\gamma}$ with $k = k(n) = o(n)$. For k we chose $k(n) = \sqrt{n}$, thus the condition $k = o(n)$ is fulfilled. Results are shown in figures A.1 and A.2. Below is a summary of simulation settings.

Table 1: Simulation settings.

Figure	Distribution	γ	Estimator	n	N	$k(n)$
A.1	Pareto	1/3	Hill	$\mathbf{n}$	2000	$\sqrt{n}$
A.2	Cauchy	1	Hill	$\mathbf{n}$	2000	$\sqrt{n}$

Both resulting graphs are constructed in the same manner. Real value of the extreme value index γ is plotted as a dashed line. Black curve represents the medians, red curve 1st quartiles and blue curve 3rd quartiles of the N sized samples of estimates. In both figures A.1 and A.2 it is evident that estimates converge to the γ as n grows. Simulations were performed using computer resources within the Aalto University School of Science "Science-IT" project.

5 Conclusions

References

[1] K. Athreya and S. Lahiri. *Measure Theory and Probability Theory*. Springer Texts in Statistics. Springer, New York, 2006.

[2] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928.

[3] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

[4] L. D. Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[5] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

[6] D. M. Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10(3):754–764, 1982.

[7] A. Rényi. On the theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungarica*, 4(3):191–231, Sep 1953.

[8] J. S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Co., Singapore, second edition edition, 2006.

A Figures

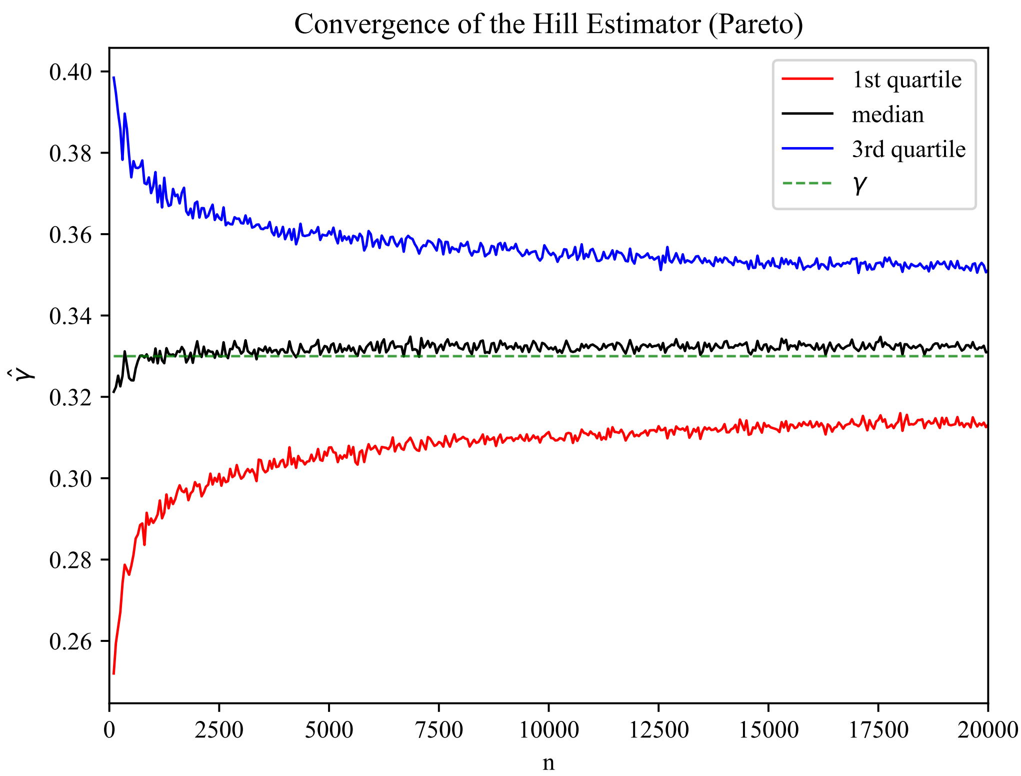

Figure A.1:

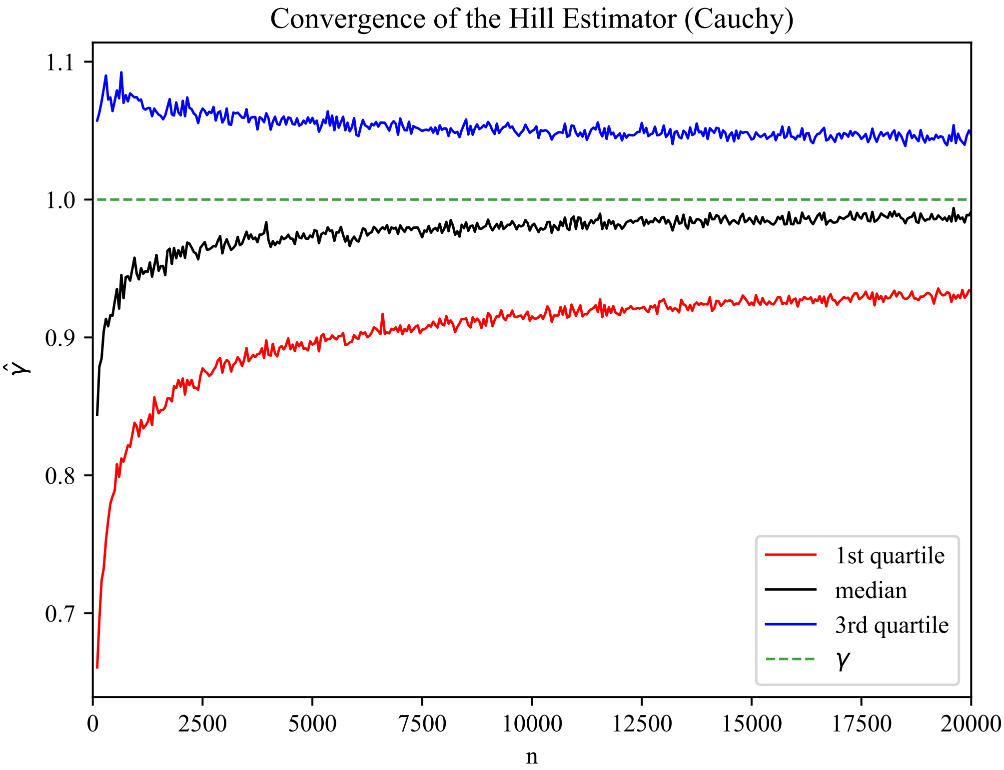

Figure A.2: testi