Thesis for the Degree of Licentiate of Technology

A Semiparametric Estimator of the Mean of Heavytailed Distributions

Joachim Johansson

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
Göteborg, Sweden 2001

A Semiparametric Estimator of the Mean of Heavytailed Distributions
JOACHIM JOHANSSON

©JOACHIM JOHANSSON, 2001

ISSN 0347-2809
Report number 2001:46
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone +46(0)31-772 1000

Chalmers University of Technology
Göteborg, Sweden 2001

Summary

An asymptotically normally distributed estimate for the expected value of a positive random variable with infinite variance is introduced. Its behavior relative to estimation using the sample mean is investigated by simulations. An example of how to apply the estimate to file-size measurements on Internet traffic is also shown.

Keywords: Pareto distribution, mean estimation, heavy tailed distributions, telecommunication

MSC2000 classification: primary – 62G32
secondary – 62F10, 62F12, 62P30

Acknowledgements

First I would like to thank professor Holger Rootzén for being an excellent supervisor, his patience in answering questions is only surpassed by his vast mathematical knowledge.

I would also like to thank my colleagues at the department of mathematical statistics at Chalmers University of Technology for making this an interesting place to work.

Finally, thanks to Dr Henrik Nyberg at Ericsson Radio Systems AB for initiating the project.

Notation

$:=$	$f := g$, f is defined as g, equivalent to $g =: f$
$X \sim F$	X has df F
$a_n \sim b_n$	$a_n/b_n \to 1$, as $n \to \infty$
$\to_{\text{a.s.}}$	convergence almost surely
$\to_d$	convergence in distribution
$\to_p$	convergence in probability
chf	characteristic function
df	distribution function
$\bar{F}(x)$	$\bar{F}(x) = \mathsf{P}(X > x)$ for some random variable $X \sim F$
GPD	generalised Pareto distribution
iid	independent and identically distributed
$\hat{M}$	an estimate of M
MDA(Φ_α)	maximum domain of attraction of the distribution Φ_α
RV_0	$L \in RV_0$ means that L is slowly varying or regularly varying with index 0
RV_α	$L \in RV_\alpha$ means that L is regularly varying with index α

Contents

1 Background

1.1 Heavy tails and what they entail

Most statistics taught at an undergraduate level, and also mostly used in practice, assume the underlying distributions to have finite variances. This property is a result of how fast the tails of the probability distribution decays. If a random variable X has distribution function $F(x)$, the variance is finite if $\int x^2 dF(x)$ converges. This, in turn, has to do with the behaviour of $\bar{F}(x) = \mathsf{P}(X > x)$ and $\mathsf{P}(X < -x)$ for large x, the **tails** of the distribution.

When the random variable X does not have finite variance, the tails of the distribution decay too slowly, meaning that there is a high probability of observing very large (positive or negative) values. Here, distributions with this property are labelled **heavy tailed**. (This term is used in various other situations as well and no universally accepted definition seem to exist). Such phenomena occur for instance when attempting to model file-size distributions on the Internet (see e.g. [4]) or in many financial and insurance applications.

The problem of analysing data with heavy tailed properties is that standard statistical methods will typically not be valid, since these are generally proven under finite variance assumptions. Lately, however, this area has received more attention, resulting in practically oriented books like e.g. [1], which presents numerous examples of applications and concrete methods in a simple and easily accessible way.

A large body of work is available on estimating quantiles of heavy tailed distributions, i.e. estimating $\mathsf{P}(X > x)$ for large x. This kind of problems occur for instance in insurance applications. If X models the size of a claim, then the risk of receiving a claim larger than some level x, is precisely $\mathsf{P}(X > x)$. The ability to estimate such quantities correctly is of crucial importance for insurance companies, when determining rates for their policies. Numerous methods for this kind of estimation problems are available, see for instance [7] for an overview and introduction to the area.

When it comes to estimating the size of the average claim, $\mathsf{E}[X]$, or the average size of a file on the Internet, the most common method is to use the sample mean. In case the distribution of X has the property that $\mathsf{P}(X > x) \approx Cx^{-\alpha}$ for large x, where $C > 0$ is a constant and $\alpha \in (1, 2)$, the sample mean will converge in distribution to a stable distribution (see Chapter 2 for a definition), not a normal as in the case with finite variance.

Correctly estimating the parameters of a stable distribution is difficult, but some very good software is available from John Nolan's home page (http://academic2.american.edu/~jpnolan). Another property of the stable distribution is that it has heavier tails than its normal counterpart, resulting in longer confidence intervals for estimates of $\mathsf{E}[X]$.

The aim of this thesis is to introduce a semi-parametric estimate of the expected value in a heavy-tailed setting. This means that a model will be suggested for the tail of the distribution, but not for the "bulk" ($\mathsf{P}(X < x)$) of it, making it reasonably flexible. The estimate will be asymptotically normal and unbiased and have an easily estimated variance. Simulations comparing it to the sample mean will also show that it exhibits a smaller median bias for finite sample sizes.

1.2 Readers road map

The thesis is divided into four chapters. In Chapter 1 a short introduction to the concept of heavy tails is given, serving to motivate the problem later addressed in Chapter 2. Chapter 2 begins by defining some concepts which will be needed and also repeats some already known results from probability theory. It then goes on to introduce a semi-parametric estimate of the expected value of a heavy-tailed distribution and examines its properties. In Chapter 3 the performance of the estimate is evaluated by simulations. In Chapter 4 the suggested estimation procedure is applied to file size measurements of TCP traffic over the Internet. Finally there are two appendices where particularly long calculations have been collected.

The hurried reader might wish to read only the first part of Chapter 2 and then skip directly to Theorem 2.1, which is the main result of the thesis. Perhaps a cursory glance at the graphs in Chapter 3, where the estimate is compared to estimation using the sample mean, will also be of interest.

2 An asymptotically normal estimate

In Section 2.1 some definitions and known results are gathered for later reference. In Section 2.2 a semi-parametric estimate of the mean of a random variable in a heavy-tailed setting is suggested and its properties are investigated in Section 2.3.

2.1 Preliminaries

2.1.1 The Landau symbols o and O

A version of the Landau symbols o and O, also called *little* o and *big* O, will be used extensively throughout. For sequences $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ and $\{R_n\}_{n=1}^{\infty}$ of random variables

$$X_n = o_p(R_n) \text{ means that } X_n = Y_n R_n \text{ and } Y_n \to_p 0$$
$$X_n = O_p(R_n) \text{ means that } X_n = Y_n R_n \text{ where } \{Y_n\}_{n=1}^{\infty} \text{ is tight.}$$

Useful properties of this notation include

$$\begin{aligned} o_p(1) + o_p(1) &= o_p(1) & (1 + o_p(1))^{-1} &= O_p(1) \\ o_p(1) + O_p(1) &= O_p(1) & o_p(R_n) &= R_n o_p(1) \\ O_p(1) o_p(1) &= o_p(1) & O_p(R_n) &= R_n O_p(1) \\ o_p(O_p(1)) &= o_p(1), \end{aligned}$$

where most are easy to prove, see for instance [21]. Further, the notation $a_n = O_+(n^\alpha)$ is introduced to denote that a_n/n^α is bounded away from zero and infinity. This is not standard notation, but will be useful later.

2.1.2 Regular variation

A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is **regularly varying** (at infinity) with index $\alpha \in \mathbb{R}$, written $f \in RV_\alpha$, if

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

For the special case $\alpha = 0$, $f \in RV_0$ is said to be **slowly varying**. It follows that if $f \in RV_\alpha$, it can be written $f(x) = x^\alpha g(x)$, where $g \in RV_0$. Regularly varying functions can be thought of as functions which behave asymptotically as power functions. For further properties, see Chapter 0 in [14] or either of [17] or [3], where the latter is very extensive.

2.1.3 Stable distributions

A random variable X is said to have a **stable distribution** if its characteristic function can be written

$$\mathsf{E}[e^{i\theta X}] = \begin{cases} \exp\{-\sigma^\alpha|\theta|^\alpha(1 - i\beta \text{sign}(\theta)\tan\frac{\pi\alpha}{2}) + i\mu\theta\} & \text{if } \alpha \neq 1, \\ \exp\{-\sigma|\theta|(1 + i\beta\frac{2}{\pi}\text{sign}(\theta)\ln|\theta|) + i\mu\theta\} & \text{if } \alpha = 1, \end{cases}$$

for some parameters $\alpha \in (0, 2]$, $\sigma \geq 0$, $\beta \in [-1, 1]$ and $\mu \in \mathbb{R}$. α is the index of stability, σ the scale, β denotes the skewness and μ the shift parameter. $\alpha = 2$ corresponds to the normal distribution.

If X follows a stable distribution, this is denoted $X \sim S_\alpha(\sigma, \beta, \mu)$. There exist other, equivalent, parametrisations of the stable distribution. For these and other properties, see the excellent book [16].

2.2 An estimate of the mean of heavytailed random variables

Let $X, X_1, X_2, \ldots, X_n$ be iid positive random variables with distribution function F, where

$$\bar{F}(x) := 1 - F(x) = cx^{-1/\xi}(1 + x^{-\delta}L(x)), \tag{2.1}$$

for $\xi \in (0, 1)$, $\delta > 0$, $L \in RV_0$ and some constant c. For $\xi \in (0, 1/2)$, X has finite variance and estimation of $\mathsf{E}[X]$ could be done using the sample mean $\bar{X} = (X_1 + \ldots + X_n)/n$ which, by the Central Limit Theorem, is asymptotically normal as $n \to \infty$. If $\xi \in (1/2, 1)$, $\bar{X}$ converges to a stable distribution (see e.g. [6]) and is still a valid estimate. There are, however, some difficulties in estimating the parameters of the limiting stable distribution. If the sample is big enough, it could be partitioned into sub-samples, for which the mean of each could be calculated. Such a procedure would render $\bar{X}_1, \ldots, \bar{X}_k$, which would be iid with a distribution that is nearly stable. Inference about the unknown parameters could then for instance be based on maximum likelihood methods. If the sample size, n, is not big enough to accommodate this method, bootstrapping, as described in [15], is an alternative.

Below, a different procedure is suggested for the model in (2.1). It will give estimators which are asymptotically normally distributed and unbiased with an easily estimated variance. In the calculations, δ will be treated as a known perturbation. The reason for this is that, normally, the sample size n would have to be very big in order to estimate it reasonably well. Instead the properties of the mean estimate will be examined as functions of δ.

Before the alternative estimate is suggested, note the following straightforward result.

Proposition 2.1
Let X be a positive random variable with distribution function F as in (2.1) and let $u_n \to \infty$ as $n \to \infty$ be a sequence of numbers. Then

$$\begin{aligned} p_n &= \mathsf{P}(X_1 > u_n) = O_+(u_n^{-1/\xi}) \\ \mu_n &= \mathsf{E}[X_1 \mathbf{1}_{\{X_1 \le u_n\}}] = O_+(u_n^{1-1/\xi}) \\ \gamma_n^2 &= \mathsf{Var}(X_1 \mathbf{1}_{\{X_1 \le u_n\}}) = O_+(u_n^{2-1/\xi}) \end{aligned}$$

Proof: $\bar{F}(x) = cx^{-1/\xi}(1 + o(1)) = x^{-1/\xi} O_+(1)$ and the results follow from standard calculations, using $\mathsf{E}[X^k] = \int_0^\infty kx^{k-1} \mathsf{P}(X > x)dx$ for positive random variables X and $k = 1, 2, 3, \ldots$. □

The standard estimate of $\mathsf{E}[X]$ is

$$\bar{X} = \int x dF_n(x) = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

where F_n is the empirical distribution function. Building on this we propose an estimate of the form

$$\hat{\mathsf{E}}[X] := \hat{M} := \hat{\mu} + \hat{\tau} := \int_0^{u_n} x dF_n(x) + \int_{u_n}^{\infty} x d\hat{F}(x), \tag{2.2}$$

where $\hat{\tau}$ is the part of $\hat{M}$ which comes from the tail of the distribution. The tail is assumed to start at some level u_n, which in the analysis will be assumed to tend to infinity.

Let $F_u(y) = \mathsf{P}(X - u_n \le y | X > u_n)$ be the distribution of the excesses over the threshold u_n. It follows from (2.1) that

$$\bar{F}_u(y) = \frac{\bar{F}(u_n + y)}{\bar{F}(u_n)} = \left(1 + \frac{y}{u_n}\right)^{-1/\xi} \frac{1 + (u_n + y)^{-\delta} L(u_n + y)}{1 + u_n^{-\delta} L(u_n)}, \tag{2.3}$$

and if $\beta_n = \beta(u_n) = u_n\xi$, then $\bar{F}_u(y)$ is a perturbed GPD, where the df of the **generalised Pareto distribution (GPD)** has the form

$$G_{\beta,\xi}(x) = \begin{cases} 1 - \left(1 + \xi\frac{x}{\beta}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - e^{-x/\beta}, & \xi = 0 \end{cases}, \quad x \in \begin{cases} [0,\infty), & \xi \geq 0 \\ [0, -\beta/\xi], & \xi < 0 \end{cases}.$$

This means that for large values of u_n, $F_u(y) \approx G_{\beta(u_n),\xi}(y)$ in the sense that

$$\lim_{u_n \uparrow y_F} \sup_{0<y<y_F-u_n} |F_u(y) - G_{\beta(u_n),\xi}(y)| = 0,$$

where y_F is the right end point of F and β is some positive function. See also Theorem 3.4.13 in [7].

By definition $\bar{F}(u_n + y) = \bar{F}(u_n)\bar{F}_u(y)$. And, for $N = N_n = |\{i : X_i > u_n\}|$, the number of X_i's which exceed u_n, we have $N \sim \text{Bin}(n, p_n = \mathsf{P}(X_1 > u_n))$, and estimation of $\bar{F}(u_n)$ may be done using

$$\widehat{\bar{F}}(u_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i > u_n\}} = \frac{N}{n}.$$

For large values of u_n, use

$$\widehat{\bar{F}}_u(y) = \bar{G}_{\hat{\beta},\hat{\xi}}(y),$$

for appropriate estimates $\hat{\xi} = \hat{\xi}_n$ and $\hat{\beta}_n = \hat{\beta}(u_n)$. Note that β will be estimated separately, i.e. $\beta = \xi u_n$ will not be used. The reason for this is to achieve greater flexibility in the parameter fitting, compensating for the underlying distribution not being an exact GPD.

We have now arrived at an alternative estimate, $\hat{M}$, of $\mathsf{E}[X]$,

$$\begin{aligned} \hat{M} &:= \int_0^{u_n} x dF_n(x) + \int_{u_n}^{\infty} x d\hat{F}(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i \mathbf{1}_{\{X_i \leq u_n\}} + \int_{u_n}^{\infty} x \frac{N}{n\hat{\beta}}\left(1 + \hat{\xi}\frac{x - u_n}{\hat{\beta}}\right)^{-1-1/\hat{\xi}} dx \qquad (2.4) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i \mathbf{1}_{\{X_i \leq u_n\}} + \hat{p}\left(u_n + \frac{\hat{\beta}}{1-\hat{\xi}}\right), \text{ for } \hat{\xi} \in (0,1), \qquad (2.5) \end{aligned}$$

where $\hat{p} = N/n$. As seen above, the main interest is in the case $\xi \in (1/2, 1)$. It will be seen later that, if $\xi \in (0,1)$ then $\mathsf{P}(\hat{\xi}_n \in (0,1)) \to 1$, so the convergence of the integral will asymptotically not be a problem.

If $\hat{\xi} \geq 1$, then $\hat{M}$ should be set to ∞ since this would indicate that the first moment of the distribution in (2.1) is infinite.

2.2.1 Estimation of β and ξ

Assume the excesses $Y_1, \ldots, Y_N$ to be iid GPD, where $Y_1 = X_1 - u_n$ and N is independent of the Y_i. [19] then gives the distribution of the conditional maximum likelihood estimates of (ξ, η), where $\eta = -\xi/\beta$. There is no closed expression for the estimates, but it turns out that they satisfy

$$\hat{\xi} = \hat{\xi}(\eta) = \frac{1}{N}\sum_{i=1}^{N} \ln(1 - \eta Y_i),$$

where

$$h(\eta) = \frac{1}{\eta} + \frac{1}{N}\Big(\frac{1}{\hat{\xi}(\eta)} + 1\Big)\sum_{i=1}^{N}\frac{Y_i}{1 - \eta Y_i} = 0$$

and $h(\eta)$, $\eta \in (-\infty, \max_i Y_i)$, is continuous at 0. These equations are then solved numerically. The case of interest here will be $\xi \in (0, 1)$, where the Cramér regularity conditions will be met, ensuring convergence of $(\hat{\beta}, \hat{\xi})$ to a bivariate normal distribution. Extensions could be made to the case $\xi > -1/2$, see [18].

2.3 Asymptotic distribution

The aim of this section is to find the asymptotic distribution of the estimate $\hat{M}$ given by (2.5). The properties of the estimates $\hat{\beta}_N$, $\hat{\xi}_N$, $\hat{p}_n$ and $\hat{\mu}_n$ are investigated first, and the results are then used for proving Theorem 2.1, which is the main result of this section.

2.3.1 Asymptotic distribution for $(\hat{\beta}, \hat{\xi})$

Extensive use will be made of the fact that if $U(\psi)$ is the score statistic for a parameter ψ and the log-likelihood function is approximately quadratic near the maximum likelihood estimate $\hat{\psi}$, then

$$\hat{\psi} - \psi \to_d \mathcal{N}(Q^{-1}b, Q^{-1}), \tag{2.6}$$

as the sample size tends to infinity. Q is the Fisher information matrix and $b = \mathsf{E}[U]$. See e.g. Chapter 5 in [11] for details.

Lemma 2.1 (GPD case)
Let $Y_1, \ldots, Y_n$ be iid GPD and $(\hat{\beta}_n, \hat{\xi}_n)$ be the maximum likelihood estimates of (β, ξ), where $\xi \in (0, 1)$. Then

$$\sqrt{n}Q_{GPD}^{1/2}\begin{pmatrix}\hat{\beta}_n - \beta \\ \hat{\xi}_n - \xi\end{pmatrix} \to_d \mathcal{N}(0, I), \quad \text{as } n \to \infty,$$

where

$$Q_{GPD} = \frac{1}{1 + 2\xi}\begin{pmatrix}\frac{1}{\beta^2} & \frac{1}{\beta(1+\xi)} \\ \frac{1}{\beta(1+\xi)} & \frac{2}{1+\xi}\end{pmatrix} \tag{2.7}$$

and I is the 2×2 identity matrix.

Proof: The log-likelihood function for one observation, y, can be written as

$$l(\beta,\xi|y) = \ln\left[\frac{1}{\beta}\Big(1+\xi\frac{y}{\beta}\Big)^{-1-1/\xi}\right],$$

and the claim follows using standard calculations and (2.6) above, see Appendix A or e.g. [19] for details. The assumption $\xi \in (0,1)$ ensures that Cramér's regularity conditions are not violated and, hence, that the maximum likelihood estimates converge to a normal distribution. □

Here it will be more interesting to look at the asymptotic distribution for $\hat{\beta}$ and $\hat{\xi}$ when the underlying model is the perturbed Pareto in Equation (2.3), the result is stated in the following lemma.

Lemma 2.2 (Perturbed case)
Let $Y_1,\dots,Y_n$ be iid with df F_u as in Equation (2.3) and with $\xi \in (0,1)$. Then

$$\sqrt{n}Q_{GPD}^{1/2}\begin{pmatrix}\hat{\beta}_n-\beta\\ \hat{\xi}_n-\xi\end{pmatrix} \to_d \mathcal{N}(0,I) \quad \text{as } n\to\infty$$

where Q_{GPD} is given by (2.7) and I is the 2×2 identity matrix. Further, for large n and u_n such that $|u_n^{-\delta}L(u_n)| < 1$,

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_n-\beta\\ \hat{\xi}_n-\xi\end{pmatrix}$$

has covariance Q^{-1}, where $Q = Q_{GPD} + Q_{Pert}$ and

$$Q_{Pert} = \begin{pmatrix}Q_{\beta\beta} & Q_{\beta\xi}\\ Q_{\beta\xi} & Q_{\xi\xi}\end{pmatrix} \quad \text{with} \tag{2.8}$$

$$\begin{aligned}
Q_{\beta\beta} &\sim -\frac{1}{\beta^2(1+2\xi)}\cdot\frac{2\delta\xi(1+\xi)}{(1+(2+\delta)\xi)}u_n^{-\delta}L(u_n)\\
Q_{\beta\xi} &\sim -\frac{1}{\beta(1+\xi)(1+2\xi)}\cdot\frac{\delta\xi(3+(6+\delta)\xi+2\xi^2)}{(1+\xi(1+\delta))(1+\xi(2+\delta))}u_n^{-\delta}L(u_n)\\
Q_{\xi\xi} &\sim \frac{2}{(1+\xi)(1+2\xi)}\cdot\frac{1}{(1+\delta\xi)(1+\xi(1+\delta))(1+\xi(2+\delta))}\Big(\delta[\delta(3+\delta)\xi^5+\\
&\quad(\delta^2+4\delta-1)\xi^4-(2\delta^2+4\delta+3)\xi^3-(\delta^2+7\delta+7)\xi^2-(5+2\delta)\xi-1]\Big)u_n^{-\delta}L(u_n),
\end{aligned}$$

and bias $Q^{-1}b$, where

$$b = \begin{pmatrix}\mathsf{E}[U_\beta]\\ \mathsf{E}[U_\xi]\end{pmatrix} \sim \frac{\delta\xi}{1+\xi(1+\delta)}\begin{pmatrix}-1/\beta\\ (\delta(\xi-1)-1)/(\xi(1+\delta\xi))\end{pmatrix}u_n^{-\delta}L(u_n) \tag{2.9}$$

Proof: See Appendix B. □

For u_n large enough to make $\|Q_{GPD}^{-1}Q_{Pert}\| < 1$ in Lemma 2.2 it is possible to use the approximation $Q^{-1} \approx Q_{GPD}^{-1} - Q_{GPD}^{-1}Q_{Pert}Q_{GPD}^{-1}$, see Appendix B for details.

2.3.2 Asymptotic distribution of $\hat{\mu}$

Lemma 2.3 (Distribution of $\hat{\mu}$)
Let $X_1, X_2, \ldots$ be positive iid random variables with df F given by (2.1). Further, let $\mu_n = \mathsf{E}[X_1 \mathbf{1}_{\{X_1 \le u_n\}}]$ and $\gamma_n^2 = \mathsf{Var}(X_1 \mathbf{1}_{\{X_1 \le u_n\}})$, with $u_n = O_+(n^{\alpha\xi})$ for some $\alpha \in (0,1)$ and $\xi \in (0,1)$. Then

$$\frac{\sqrt{n}}{\gamma_n}(\hat{\mu} - \mu_n) \to_d \mathcal{N}(0,1), \quad \text{as } n \to \infty,$$

or, equivalently,

$$\mathsf{E}[\exp\{it\frac{\sqrt{n}}{\gamma_n}(\hat{\mu} - \mu_n)\}] \to e^{-t^2/2}, \quad \text{as } n \to \infty,$$

where $\hat{\mu}$ is defined in Equation (2.2).

Proof: The proof is straightforward in that all that has to be done is to verify that the Lindeberg-Feller Theorem (see e.g. Chapter 2 in [6]) is applicable. To this end, note that

$$\frac{\sqrt{n}}{\gamma_n}(\hat{\mu} - \mu_n) = \sum_{k=1}^{n} \frac{X_k \mathbf{1}_{\{X_k \le u_n\}} - \mu_n}{\gamma_n \sqrt{n}} =: \sum_{k=1}^{n} Y_{k,n},$$

where $\mathsf{E}[Y_{k,n}] = 0$, $\mathsf{E}[Y_{k,n}^2] = 1/n$ and hence $\sum_{k=1}^{n} \mathsf{E}[Y_{k,n}] = 1$ for all $n \ge 1$.

Further, by Proposition 2.1,

$$\mu_n \pm \epsilon\gamma_n\sqrt{n} = \pm n^{\alpha(\xi - 1/2) + 1/2} O_+(1) \to \pm\infty, \quad \text{as } n \to \infty$$

for all $\alpha \in (0,1)$, $\xi \in (0,1)$ and $\epsilon > 0$, where $u_n = O_+(n^{\alpha\xi})$ was used. This means that

$$\begin{aligned}
\sum_{k=1}^{n} \mathsf{E}[|Y_{k,n}|^2; |Y_{k,n}| > \epsilon] &= \frac{1}{\gamma_n^2} \mathsf{E}[(X_1 \mathbf{1}_{\{X_1 \le u_n\}} - \mu_n)^2; |X_1 \mathbf{1}_{\{X_1 \le u_n\}} - \mu_n| > \epsilon\gamma_n\sqrt{n}] \\
&\le \frac{u_n^2}{\gamma_n^2} \mathsf{P}(|X_1 \mathbf{1}_{\{X_1 \le u_n\}} - \mu_n| > \epsilon\gamma_n\sqrt{n}) \\
&= \frac{u_n^2}{\gamma_n^2} \Big(\mathsf{P}(X_1 \mathbf{1}_{\{X_1 \le u_n\}} > \mu_n + \epsilon\gamma_n\sqrt{n}) \\
&\quad + \mathsf{P}(X_1 \mathbf{1}_{\{X_1 \le u_n\}} < \mu_n - \epsilon\gamma_n\sqrt{n}) \Big) \\
&= \frac{u_n^2}{\gamma_n^2} (\bar{F}(\mu_n + \epsilon\gamma_n\sqrt{n}) + F(\mu_n - \epsilon\gamma_n\sqrt{n})) \to 0, \quad \text{as } n \to \infty,
\end{aligned}$$

since $F(x) = 0$ for $x < 0$, $\bar{F}(x) = x^{-1/\xi} O_+(1)$ and $u_n^2/\gamma_n^2 = n^{\alpha} O_+(1)$. The convergence holds for all $\epsilon > 0$, $\alpha \in (0,1)$ and $\xi \in (0,1)$. The result now follows from the Lindeberg-Feller Theorem. □

2.3.3 Joint distribution of the parameters

Lemma 2.4 (Joint distribution)
Let $X_1, X_2, \dots$ be positive iid random variables with distribution function F as in (2.1) and $N = |\{X_i : X_i > u_n\}| \sim \text{Bin}(n, p_n)$, where $p_n = \bar{F}(u_n)$ and $u_n = O_+(n^{\alpha\xi})$ for some $\alpha \in (0,1)$ and $\xi \in (0,1)$. Then

$$\phi(t_1,t_2,t_3,t_4) = \mathsf{E}[\exp\{it_1\frac{\sqrt{n}}{\gamma_n}(\hat{\mu}-\mu_n) + i\sqrt{np_n}Q_{GPD}^{1/2}(t_2,t_3)\begin{pmatrix}\hat{\beta}_N-\beta\\ \hat{\xi}_N-\xi\end{pmatrix} + it_4\frac{\sqrt{n}(\hat{p}-p_n)}{\sqrt{p_n(1-p_n)}}\}]$$
$$\to \exp\{-\frac{t_1^2}{2} - \frac{1}{2}(t_2,t_3)\begin{pmatrix}t_2\\ t_3\end{pmatrix} - \frac{t_4^2}{2}\} \quad \text{as } n\to\infty,$$

where Q_{GPD} is given by (2.7), $\mu_n = \mathsf{E}[X_1\mathbf{1}_{\{X_1\le u_n\}}]$, $\gamma_n^2 = \mathsf{Var}(X_1\mathbf{1}_{\{X_1\le u_n\}})$ and $\hat{p} = N/n$.

Proof: Let

$$\phi_{\mu|N}(t_1) = \mathsf{E}[\exp\{it_1\frac{\sqrt{n}}{\gamma_n}(\hat{\mu}-\mu_n)\}|N]$$
$$\phi_{\beta,\xi|N}(t_2,t_3) = \mathsf{E}[\exp\{i\sqrt{np_n}Q_{GPD}^{1/2}(t_2,t_3)\begin{pmatrix}\hat{\beta}_N-\beta\\ \hat{\xi}_N-\xi\end{pmatrix}\}|N]$$
$$\phi_{p|N}(t_4) = \mathsf{E}[\exp\{it_4\frac{\sqrt{n}(\hat{p}-p_n)}{\sqrt{p_n(1-p_n)}}\}|N] = \exp\{it_4\frac{\sqrt{n}(\hat{p}-p_n)}{\sqrt{p_n(1-p_n)}}\},$$

then $\phi(t_1,t_2,t_3,t_4) = \mathsf{E}[\phi_{\mu|N}(t_1)\phi_{\beta,\xi|N}(t_2,t_3)\phi_{p|N}(t_4)]$. Note that, conditional on N, $\hat{\mu}$ is independent of $(\hat{\beta}_N, \hat{\xi}_N)$.

If $\{\phi_n\}_{n=1}^{\infty}$ is a sequence of characteristic functions such that $\phi_n(t) \to \phi(t)$, then there is a constant n_0 such that, for each $n > n_0$ there is a $\delta > 0$ such that $|\phi_n(t) - \phi(t)| < \delta$. This means that

$$\begin{aligned}\mathsf{P}(|\phi_N(t)-\phi(t)|<\delta) &= \mathsf{P}(|\phi_N(t)-\phi(t)|<\delta, N>n_0)\\ &\quad+\mathsf{P}(|\phi_N(t)-\phi(t)|<\delta, N\le n_0)\\ &= \mathsf{P}(N>n_0) + \mathsf{P}(|\phi_N(t)-\phi(t)|<\delta, N\le n_0) \to 1, \quad \text{as } n\to\infty.\end{aligned}$$

Using Lemmas 2.2 and 2.3 together with the above property, means that

$$\phi_{\mu|N}(t_1) \to_p \exp\{-\frac{t_1^2}{2}\} \quad \text{and}$$
$$\phi_{\beta,\xi|N}(t_2,t_3) \to_p \exp\{-\frac{1}{2}(t_2,t_3)\begin{pmatrix}t_2\\ t_3\end{pmatrix}\},$$

Naturally, the product $\phi_{\mu|N}(t_1)\phi_{\beta,\xi|N}(t_2,t_3)$ will also converge in probability.

Further, since $N \sim \text{Bin}(n, p_n)$, $\hat{p} = N/n$ and $np_n = n^{1-\alpha}O_+(1)$, it follows from the Lindeberg-Feller Theorem that

$$\frac{\sqrt{n}(\hat{p} - p_n)}{\sqrt{p_n(1-p_n)}} \to_d \mathcal{N}(0,1) \quad \text{and} \quad \phi_{p|N}(t_4) \to_d \exp\{it_4\mathcal{N}(0,1)\}.$$

Finally, then

$$\phi_{\mu|N}(t_1)\phi_{\beta,\xi|N}(t_2,t_3)\phi_{p|N}(t_4) \to_d \exp\{-\frac{t_1^2}{2}\}\exp\{-\frac{1}{2}(t_2,t_3)\begin{pmatrix}t_2\\t_3\end{pmatrix}\}\exp\{it_4\mathcal{N}(0,1)\},$$

and the claim follows by taking expectations. □

All the tools are now in place for proving the main theorem of this section.

Theorem 2.1 (Distribution of $\hat{M}$)
Let $X_1, \dots, X_n$ be positive iid random variables with distribution function F, such that $\bar{F}(x) = cx^{-1/\xi}(1 + x^{-\delta}L(x))$, for some constants $c > 0$, $\xi \in (0,1)$ and $\delta > 0$, where $L(x)$ is a slowly varying function. With $u_n = O_+(n^{\alpha\xi})$ for some $\alpha \in (0,1)$, $p_n = \mathsf{P}(X_1 > u_n)$, $\mu_n = \mathsf{E}[X_1\mathbf{1}_{\{X_1 \le u_n\}}]$, $\gamma_n^2 = \mathsf{Var}(X_1\mathbf{1}_{\{X_1\le u_n\}})$, β as in (2.3), $M = \mathsf{E}[X_1]$ and $\hat{M}$ as in (2.5),

$$\frac{\sqrt{n}}{\gamma_n\sqrt{k_n}}(\hat{M} - M) \to_d \mathcal{N}(0,1),$$

where

$$k_n = 1 + \frac{p_n(1-p_n)}{\gamma_n^2}\Big(u_n + \frac{\beta}{1-\xi}\Big)^2 + \frac{p_n\beta^2}{\gamma_n^2}\frac{(1+\xi)^2}{(1-\xi)^4} = O_+(1).$$

Proof: First note that Equation (2.3) states that $\beta = O_+(u_n)$. From (2.5) follows that

$$\begin{aligned}
\frac{\sqrt{n}}{\gamma_n}(\hat{M} - M) &= \frac{\sqrt{n}}{\gamma_n}(\hat{\mu} - \mu_n) + \frac{\sqrt{n}}{\gamma_n}(\hat{p} - p_n)\Big(u_n + \frac{\hat{\beta}_N}{1-\hat{\xi}_N}\Big) + \frac{\sqrt{n}}{\gamma_n}\Big(\frac{\hat{\beta}_N}{1-\hat{\xi}_N} - \frac{\beta}{1-\xi}\Big) \\
&= \{\text{Taylor expand and isolate all troublesome } \beta\text{-terms}\} \\
&= \frac{\sqrt{n}}{\gamma_n}(\hat{\mu} - \mu_n) + \frac{\sqrt{n}}{\gamma_n}(\hat{p} - p_n)\Big(u_n + \frac{\beta}{1-\xi} + \frac{1}{1-\xi}(\hat{\beta}_N - \beta) \\
&\quad + \beta\sum_{k=1}^{\infty} k\frac{(\hat{\xi}_N - \xi)^k}{(1-\xi)^{k+1}} + O_p((\hat{\beta}_N - \beta)^2 + (\hat{\xi}_N - \xi)^2)\Big) \\
&\quad + \frac{\sqrt{n}}{\gamma_n}p_n\Big(\frac{1}{1-\xi}(\hat{\beta}_N - \beta) + \frac{\beta}{(1-\xi)^2}(\hat{\xi}_N - \xi) \\
&\quad + \beta\sum_{k=1}^{\infty} k\frac{(\hat{\xi}_N - \xi)^k}{(1-\xi)^{k+1}} + O_p((\hat{\beta}_N - \beta)^2 + (\hat{\xi}_N - \xi)^2)\Big)
\end{aligned}$$

$$= \left\{ \begin{array}{l} \text{Use Proposition 2.1 together with the Continuous} \\ \text{Mapping Theorem and } Q_{GPD}^{-1} \text{ from Lemma 2.1} \end{array} \right\}$$

$$= \underbrace{\frac{\sqrt{n}}{\gamma_n}(\hat{\mu} - \mu_n)}_{\to_d \mathcal{N}(0,1)} + \frac{\sqrt{p_n(1-p_n)}}{\gamma_n}\left(u_n + \frac{\beta}{1-\xi}\right) \underbrace{\frac{\sqrt{n}}{\sqrt{p_n(1-p_n)}}(\hat{p} - p_n)}_{\to_d \mathcal{N}(0,1)}$$

$$+ \frac{\sqrt{p_n}}{\gamma_n}\frac{\beta}{(1-\xi)^2}\underbrace{\sqrt{np_n}(\hat{\xi}_N - \xi)}_{\to_d N(0,(1+\xi)^2)} + o_p(1).$$

The result now follows by using Lemma 2.4 and the Continuous Mapping Theorem (see e.g. Theorem 29.2 in [2]). The sequence

$$k_n = 1 + \frac{p_n(1-p_n)}{\gamma_n^2}\left(u_n + \frac{\beta}{1-\xi}\right)^2 + \frac{p_n\beta^2}{\gamma_n^2}\frac{(1+\xi)^2}{(1-\xi)^4} = O_+(1)$$

by Proposition 2.1. □

Note that there is no closed expression for the asymptotic variance of the estimate $\hat{M}$ under the conditions in Theorem 2.1. This is because only the tail behaviour of the distribution function F is specified and hence the assumptions do not determine the value of γ_n^2. An exact limit could be obtained, should F be more closely specified, but this would in turn mean placing extra conditions on $L(x)$. Since $L(x)$ is not known in practice, such conditions would be of limited practical use.

3 Performance of the estimate

Two methods for estimation of the expected value of a heavy-tailed random variable which satisfies (2.1), are compared. The standard way, estimation using the empirical mean, is compared to the estimate (2.5). This is done in Section 3.1. In Section 3.2, the two methods are compared by simulation for some selected cases.

3.1 Comparison to the empirical mean

If X is a positive random variable with distribution as in (2.1) and with $\xi \in (1/2, 1)$, then it has infinite variance but finite mean, $\mathsf{E}[X] = M$. The usual way of estimating M would be to use something like Theorem 7.7 in Chapter 2 in [6]:

Theorem 3.1
Suppose $X_1, X_2, \ldots$ are iid positive random variables with a distribution that satisfies
$\mathsf{P}(|X_1| > x) = x^{-\nu} L(x)$, where $\nu < 2$ and $L \in RV_0$. Let $\bar{X}_n = (X_1 + \ldots + X_n)/n$,

$$a_n = n^{-1} \inf\{x : \mathsf{P}(|X_1| > x) \leq n^{-1}\} \quad \text{and} \quad b_n = \mathsf{E}[X_1 \mathbf{1}_{\{|X_1| \leq a_n\}}].$$

Then $(\bar{X}_n - b_n)/a_n \to_d Y$ as $n \to \infty$, where Y is a stable random variable, $Y \sim S_\nu(\sigma, \kappa, M)$ for some parameters $\sigma \geq 0$ and $\kappa \in [-1, 1]$, where $M = \mathsf{E}[X_1]$ if $\nu > 1$. □

Based on Theorem 3.1 above, a confidence interval for M would look something like

$$\bar{X}_n - s_1 a_n \leq M \leq \bar{X}_n - s_0 a_n \qquad (1 - \gamma),$$

where $(1-\gamma)$ is the approximate confidence level and s_0 and s_1 are quantiles of the stable distribution. The interesting "first order" parameter here is really a_n which determines the convergence rate for the estimate. Under the assumption that the df F is as in Equation 2.1, $a_n \sim (constant) \cdot n^{\xi-1}$.

On the other hand, the estimate in Theorem 2.1 leads to a confidence interval

$$\hat{M} - z_{\gamma/2}\sqrt{k_n}\frac{\gamma_n}{\sqrt{n}} \leq M \leq \hat{M} + z_{\gamma/2}\sqrt{k_n}\frac{\gamma_n}{\sqrt{n}} \qquad (1-\gamma), \tag{3.1}$$

where $z_{\gamma/2}$ is a suitable normal distribution fractile, $(1-\gamma)$ the approximate confidence level and k_n is defined in Theorem 2.1. As above, the interesting term to look at is $\sqrt{k_n}\gamma_n/\sqrt{n}$. It follows from Proposition 2.1 that

$$\frac{\sqrt{k_n}\gamma_n}{\sqrt{n}} = n^{\alpha(\xi-\frac{1}{2})-\frac{1}{2}}O_+(1), \quad \text{for some } \alpha \in (0,1).$$

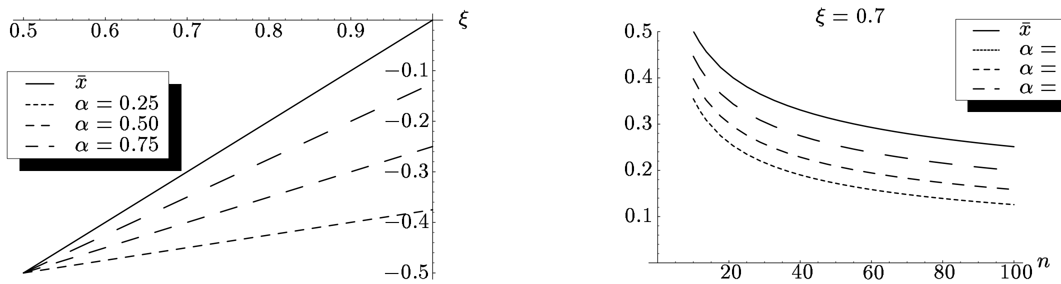

Figure 3.1: *Left:* The approximate behaviour of the logarithm of the confidence interval width. (If the width is proportional to n^ρ, the graph displays ρ, as a function of ξ and $\alpha \in (0,1)$). *Right:* The width of the confidence intervals as functions of the number of observations, n, for a fixed value $\xi = 0.7$ and for a selection of α's. $\bar{x}$ corresponds to estimation based on the sample mean, and the α's to estimation based on Equation 2.5 using threshold $u_n = n^{\alpha\xi}$.

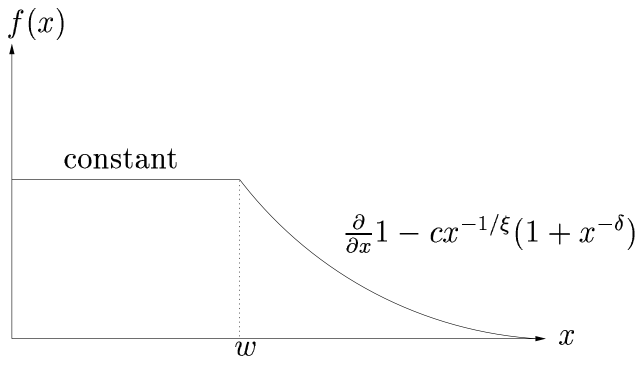

Figure 3.2: An example of how the density might look.

Figure 3.1 suggests that the optimal parameter selection would be $\alpha = 0$, corresponding to using all observations for the tail estimation. This goes against intuition, for consider the a density such as the one in Figure 3.2. In this case it would be natural to select α in such a way as to make $u_n = w$, where w denotes the starting-point of the tail. The reason Figure 3.1 suggests $\alpha = 0$ is a erroneous extrapolation of the tail-behaviour of the distribution $F(x)$ in Equation (2.1) to be valid for small x. Still, what the figure does show is that the estimate $\hat{M}$ will converge faster than $\bar{X}_n$, regardless of $\alpha \in (0, 1)$. It does not, however, say anything about bias.

Calculation of $\mathsf{Bias}(\hat{M})$ for finite sample sizes is difficult without imposing extra conditions on the distribution function F in (2.1). Instead of performing such calculations, simulations are made in the next section for a few special cases.

3.2 A simulation study

Two simulations were made in order to determine the behaviour of $\hat{M}$ itself, and its relation to $\bar{X}_n$. The parameters β and ξ were estimated using the Splus program-package EVIS, Version 3, written by Alexander J McNeil. Refinements could be made for these estimates, of course, and there is a large body of work done in this area, see for instance [19] or [20] and the references therein. Improvements to $\bar{X}_n$ could also be made, for instance by bootstrapping from the sample as described in [15].

Depending on the value of $\hat{\xi}$, different estimates of the mean are made, as illustrated in Figure 3.3. $\hat{\xi} > 1$ would indicate that the mean is infinite, and so the estimate should be infinity in this case. Further, if $\hat{\xi} < 0$ this would indicate a distribution with finite tail and then the ordinary Central Limit Theorem will be used.

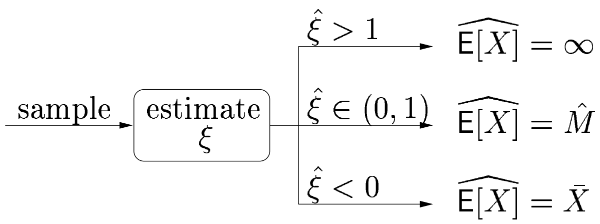

Figure 3.3: The figure shows how $\hat{M}$ is used in the simulations.

3.2.1 Exact Pareto

In the first simulation, 500 samples, each of size 1000, were drawn from a generalised Pareto distribution with parameters $\beta = 10$ and $\xi = 0.7$, corresponding to an expected value of 33.3. The estimation strategy in Figure 3.3 was compared to the one using only $\widehat{\mathsf{E}[X]} = \bar{X}$. The results are displayed in Figure 3.4.

It may be noted that the trend in the $\hat{\beta}$ graph is theoretically justified since, if Y is GPD distributed with some parameters β and ξ, then

$$\mathsf{P}(Y > u+v) = \frac{\mathsf{P}(Y > u+v|Y > u)}{\mathsf{P}(Y > u)} = \frac{\left(1+\xi\frac{u+v}{\beta}\right)^{-1/\xi}}{\left(1+\xi\frac{u}{\beta}\right)^{-1/\xi}} = \left(1+\xi\frac{v}{\beta+\xi u}\right)^{-1/\xi},$$

hence the estimate of β is expected to grow linearly with the threshold u_n. The increased variability in $\hat{\beta}$ and $\hat{\xi}$ for larger u_n is due to the estimates being based on fewer observations as the threshold increases.

A larger median bias for $\bar{X}_n$, compared to $\hat{M}$, is noted. This is an expected result since the underlying model is Pareto, and hence perfectly adapted to $\hat{M}$.

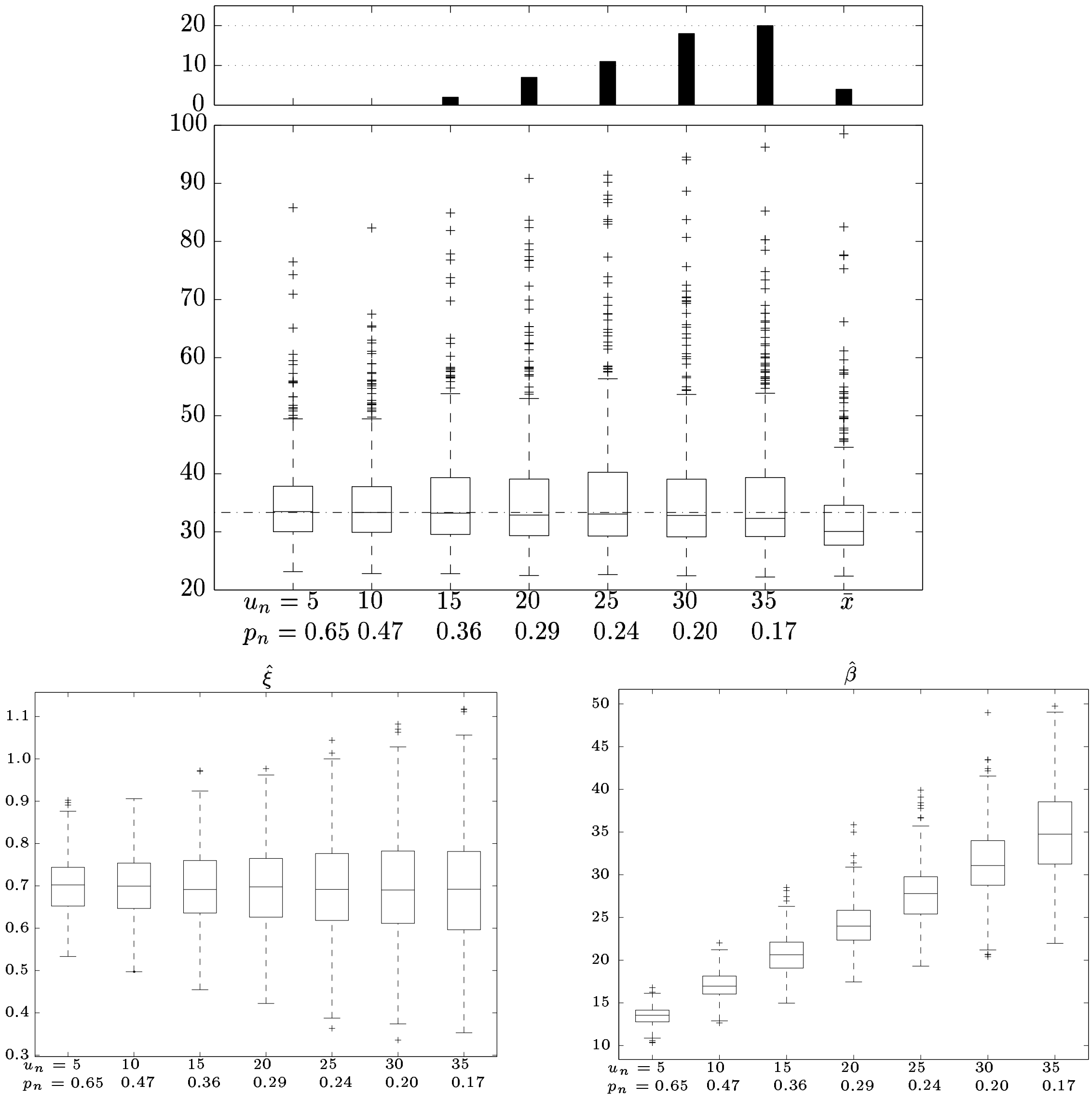

Figure 3.4: Results from the Pareto simulation. *Top:* A boxplot showing the results of the estimation strategy in Figure 3.3 compared to using $\bar{X}$. The dash-dotted line shows the true expected value. Above the boxplot is a graph showing the number of outliers not displayed. For u_n between 25 and 35 the expected value was estimated to ∞ 2, 6 and 9 times, respectively. The largest outlier for $\bar{X}_n$ was at 389 which was larger than the ones for u_n between 5 and 20, but smaller than for the other thresholds. *Bottom left:* Estimates of ξ for different values of u_n. *Bottom right:* Estimates of β for different values of u_n.

3.2.2 Distribution with perturbed Pareto tail

The above simulation with an exact underlying Pareto distribution is of course favourable for $\hat{M}$, since the correct model is assumed for the data, whereas no such assumptions are made for $\bar{X}_n$. We now study the behaviour of the estimates for a more general case.

A few simulations were done where samples were drawn from the pdf in Figure 3.2 with parameters $w = 10$, $\xi = 0.7$ and $\delta = 1$, corresponding to a mean of approximately 15.7. 500 samples, varying in size from 100, up to 100000 were generated and the estimates calculated for different values of the threshold u_n. The results are plotted in Figures 3.5 – 3.8.

For $u_n = 5$ the estimate is very biased. This is as expected since the density, $f(x)$, is constant up to $x = 10$ and then decreases polynomially.

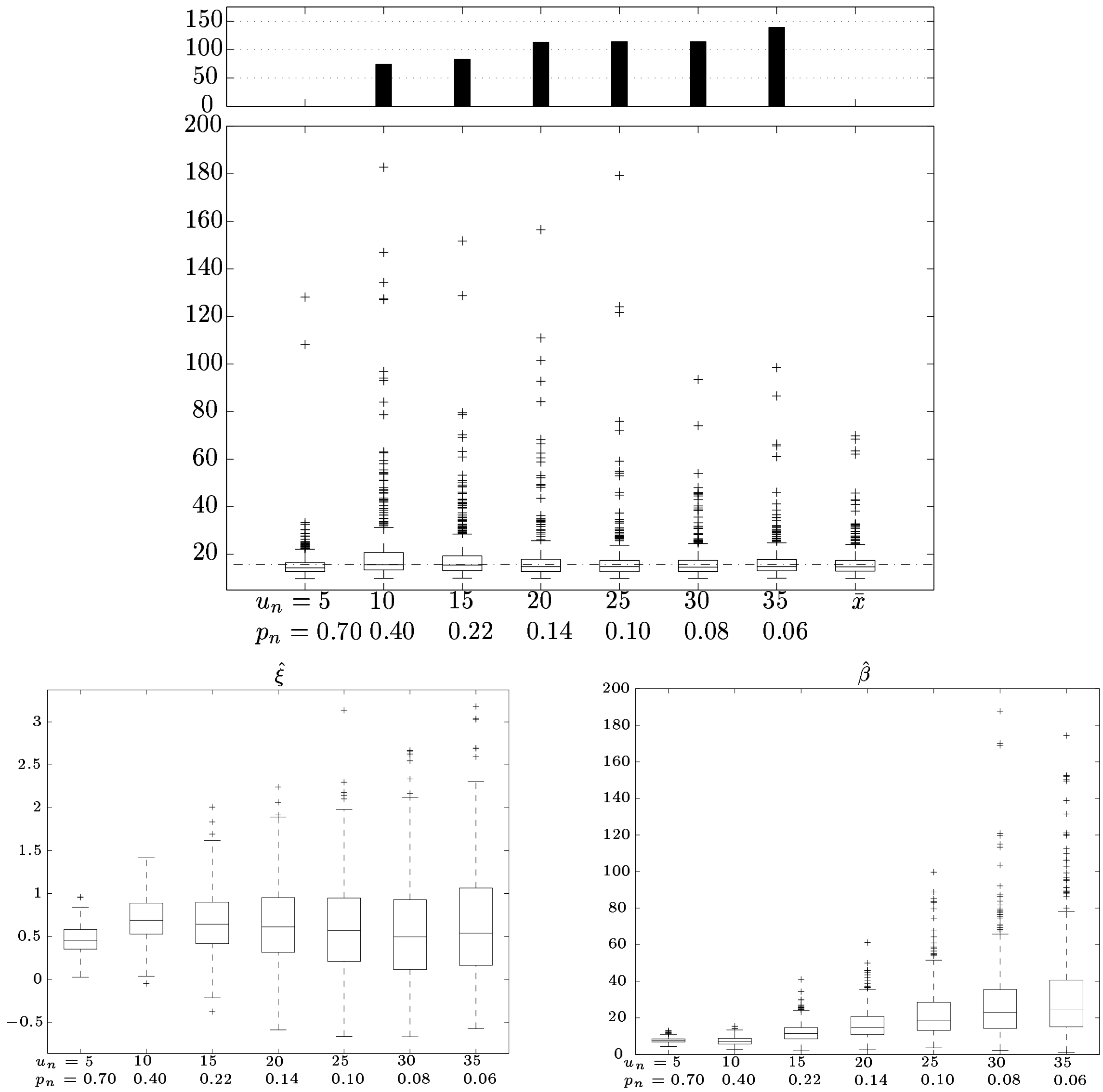

Figure 3.5: Results from the simulation with perturbed Pareto tail. The graphs are based on 500 samples, each of size 100. *Top:* A boxplot showing the results of the estimation strategy in Figure 3.3 compared to using $\bar{X}_n$. The dash-dotted line shows the true expected value. Above the boxplot is a graph showing the number of outliers not displayed. For u_n between 5 and 35 the expected value was estimated to ∞ 66, 83, 109, 113, 113 and 137 times, respectively for u_n between 10 and 35. The largest outlier for $\bar{X}_n$ was at 70 which was smaller than for the other thresholds. *Bottom left:* Estimates of ξ for different values of u_n. *Bottom right:* Estimates of β for different values of u_n.

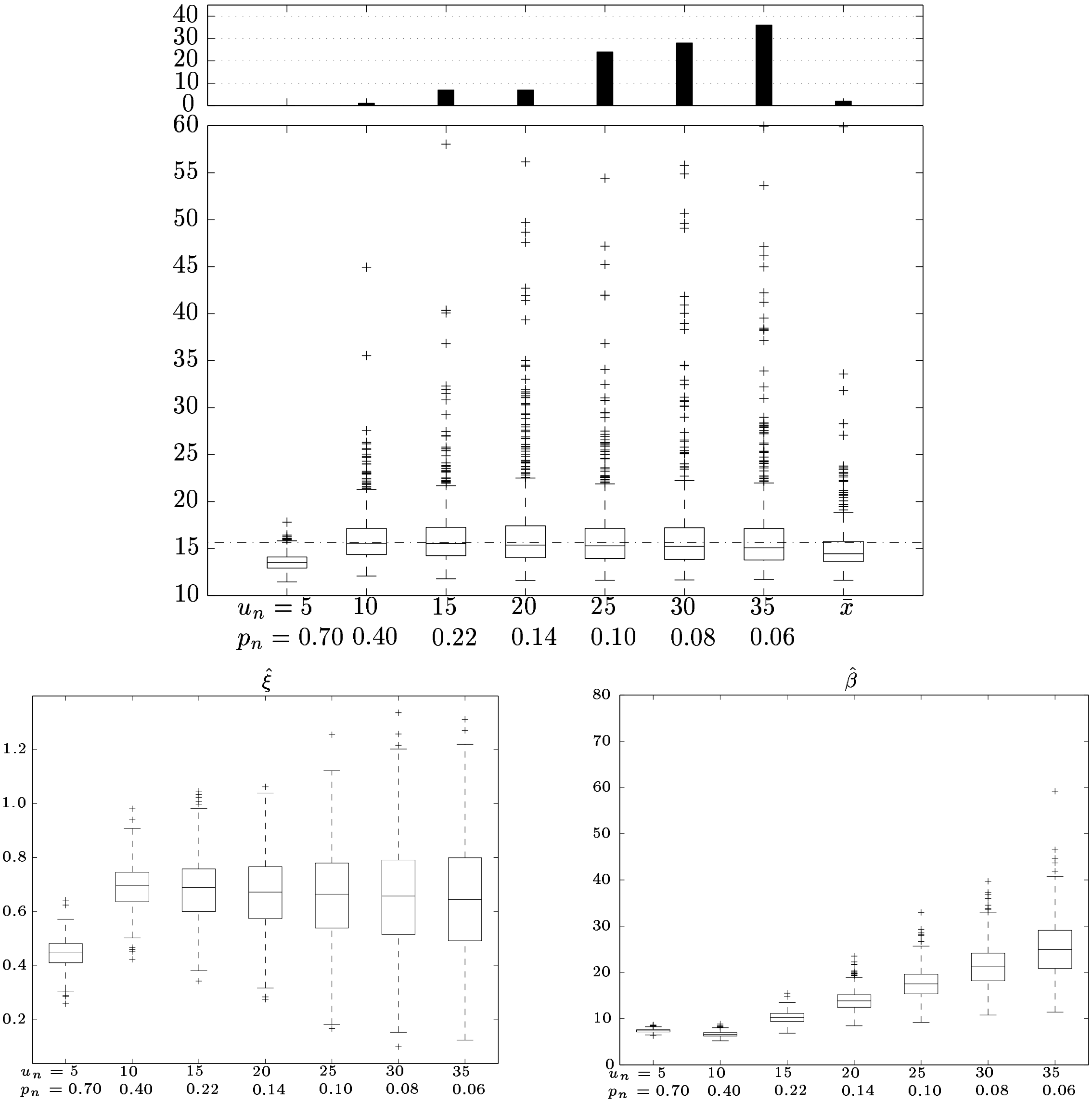

Figure 3.6: Results from the simulation with perturbed Pareto tail. The graphs are based on 500 samples, each of size 1000. *Top:* A boxplot showing the results of the estimation strategy in Figure 3.3 compared to using $\bar{X}_n$. The dash-dotted line shows the true expected value. Above the boxplot is a graph showing the number of outliers not displayed. For u_n between 15 and 35 the expected value was estimated to ∞ 4, 5, 17, 21 and 31 times, respectively, for u_n between 15 and 35. The largest outlier for $\bar{X}_n$ was at 89 which was larger than the ones for u_n equal to 5, but smaller than for the other thresholds. *Bottom left:* Estimates of ξ for different values of u_n. *Bottom right:* Estimates of β for different values of u_n.

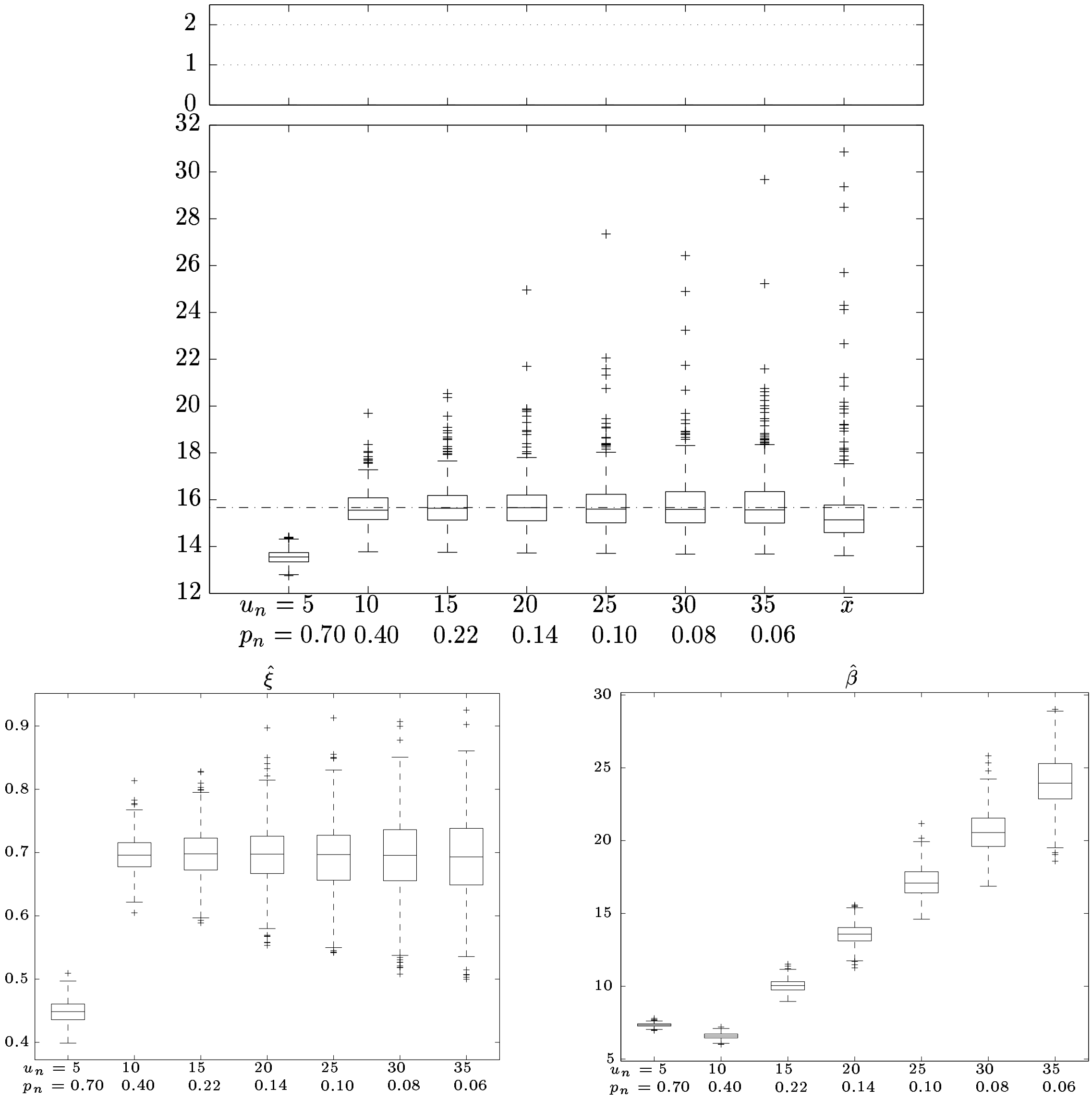

Figure 3.7: Results from the simulation with perturbed Pareto tail. The graphs are based on 500 samples, each of size 10000. *Top:* A boxplot showing the results of the estimation strategy in Figure 3.3 compared to using $\bar{X}_n$. The dash-dotted line shows the true expected value. Above the boxplot is a graph showing the number of outliers not displayed. For u_n between 5 and 35 the expected value never estimated to ∞. The largest outlier for $\bar{X}_n$ was at 31 which was larger than for all thresholds for $\hat{M}$. *Bottom left:* Estimates of ξ for different values of u_n. *Bottom right:* Estimates of β for different values of u_n.

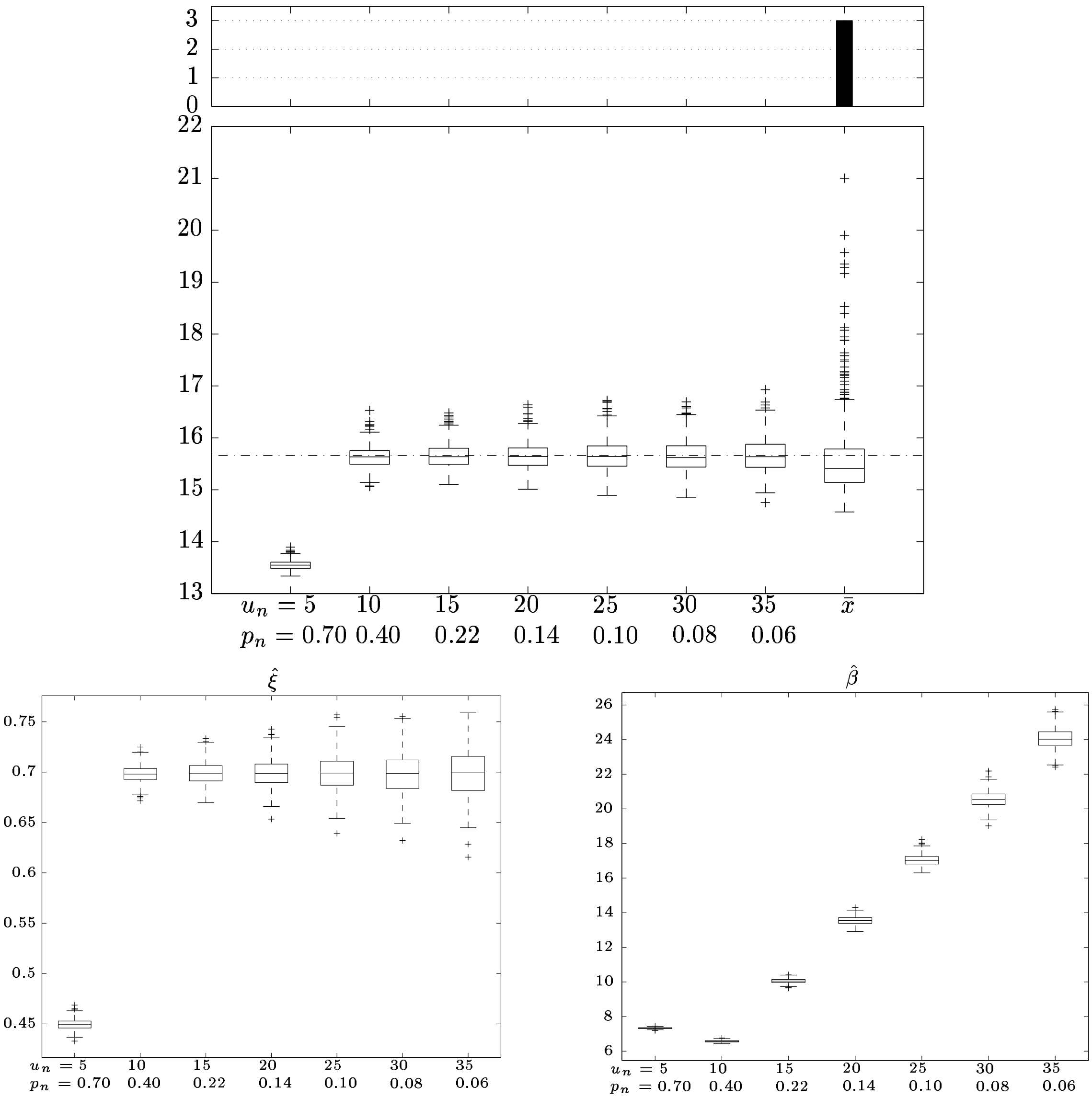

Figure 3.8: Results from the simulation with perturbed Pareto tail. The graphs are based on 500 samples, each of size 100000. *Top:* A boxplot showing the results of the estimation strategy in Figure 3.3 compared to using $\bar{X}_n$. The dash-dotted line shows the true expected value. Above the boxplot is a graph showing the number of outliers not displayed. The largest outlier for $\bar{X}_n$ was at 33 which was larger than for the $\hat{M}$-estimates. *Bottom left:* Estimates of ξ for different values of u_n. *Bottom right:* Estimates of β for different values of u_n.

For comparison with the earlier results, a similar simulation was made for a finite variance case, using parameters $\xi = 0.3$, $\delta = 1$ and $w = 10$. 500 samples, each of size 10000, were generated and the mean was estimated using the strategy in Figure 3.3. The result is displayed in Figure 3.9. For this case, no apparent difference is noted between $\hat{M}$ and $\bar{X}_n$, both appear to be unbiased and have the same variance. Surprisingly, this is also the case for $\hat{M}$ with $u_n = 5$, where normally a bias would have been expected. An explanation for this is that the tails are very light and their influence on the estimate is small.

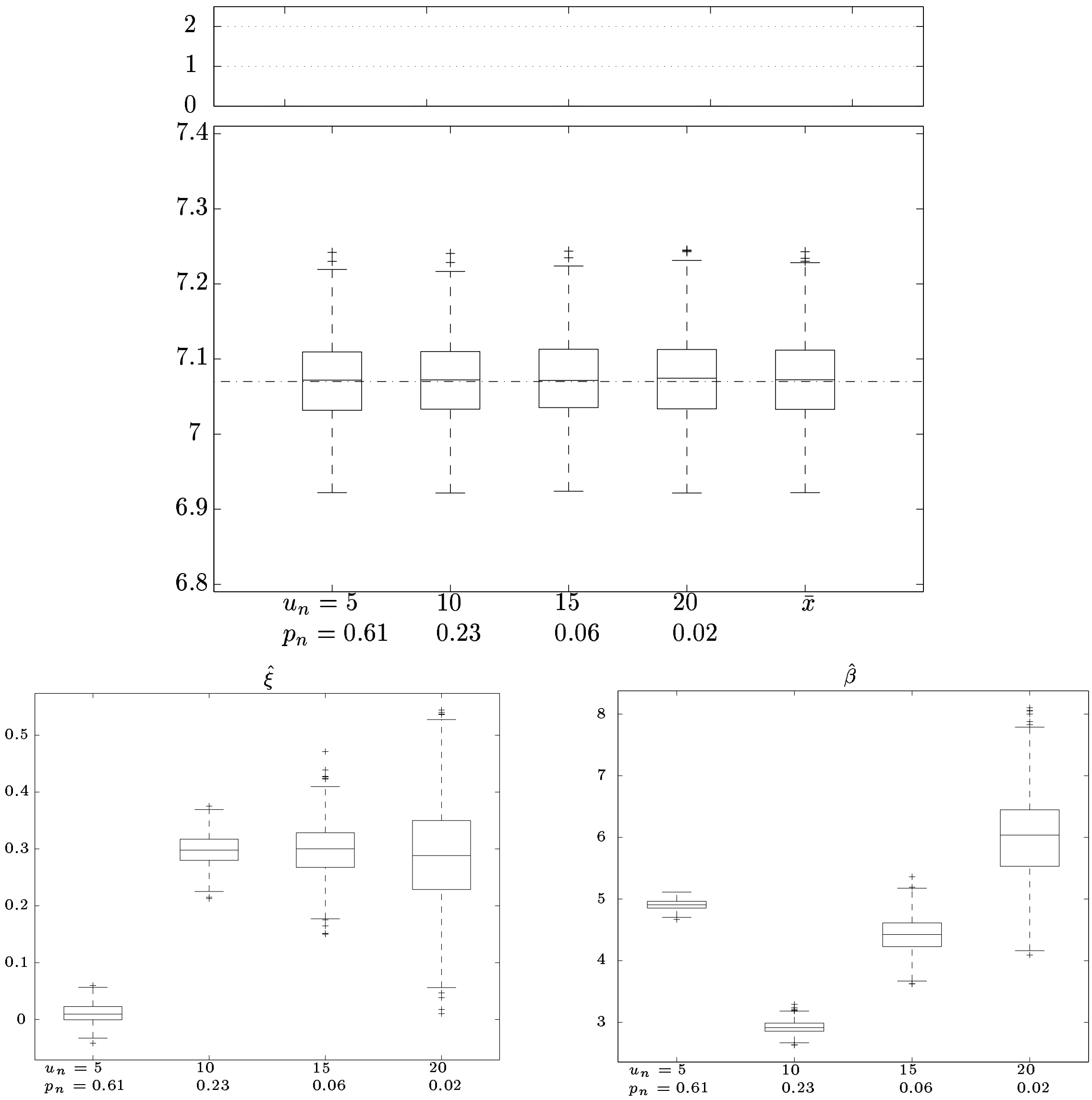

Figure 3.9: Results from the simulation with perturbed Pareto tail. The graphs are based on 500 samples, each of size 10000. *Top:* A boxplot showing the results of the estimation strategy in Figure 3.3 compared to using $\bar{X}_n$. The dash-dotted line shows the true expected value. Above the boxplot is a graph showing the number of outliers not displayed. *Bottom left:* Estimates of ξ for different values of u_n. *Bottom right:* Estimates of β for different values of u_n.

4 Applications

The aim of this chapter is to outline applications of $\hat{M}$ to telecommunication data. Focus will be on the estimate, rather than on how to model network traffic, which is a large area on its own. References to this area will therefore be very sketchy and we will not go into any details, but see for instance [9], [12], [8], [10] and its companion paper [13], for suggestions of different models.

When attempting to describe the behaviour of Internet-traffic data one often arrives at a model like

$$X(t) =_d \mu + \text{noise},$$

where $X(t)$ describes the traffic intensity into a network node at time t, μ is a constant mean traffic intensity and the noise could for instance be fractional Brownian noise. This model will be interesting when analysing network behaviour and perhaps augmenting the network.

Sometimes it will be more interesting to look at the accumulated traffic into a node

$$A(t) = \int_0^t X(s)ds =_d \nu t + \text{noise},$$

where ν is a constant and noise could be fractional Brownian motion. Such a process is interesting when it comes to bandwidth and buffer dimensioning in the network. There are various models for the processes X and A, but most of them have in common that μ or ν are expected values of some random variable. Hence, in order to have some idea of how X or A behaves, this mean will have to be estimated. In the heavy-tailed setting discussed earlier, $\hat{M}$ could be used for this purpose.

4.1 HTTP data

The data examined is the file sizes in HTTP traffic from a modem pool. It was generously supplied by Dr Attila Vidács at the High Speed Networks Laboratories at the Technical University of Budapest, and is displayed in Figure 4.1 below.

Note that there are six large observations, five of which appear closely together. A more detailed look at the data shows that five of these were generated by the same user and that four appeared back to back in time and originated from the same web site. This kind of behaviour might be observed for instance when the HTTP protocol is used for file transfers instead of using FTP. It is not clear that this was the case, but in order to keep the model simple, these six large observations were taken out when the mean was estimated, so that the Pareto-tail assumption would more realistic, see Figures 4.2 and 4.3.

The mean was then estimated using the same strategy as in Figure 3.3. The result is displayed in Figure 4.4 below. Visual inspection shows that the mean might be around $1.25 \cdot 10^4$. Estimation using the sample mean, $\bar{X}_n$, without removing the six largest observations resulted in an estimate of about $1.5 \cdot 10^4$, and when the observations were removed this number changed to slightly below $1.2 \cdot 10^4$.

Two observations could be made. First, estimation based on the sample mean is sensitive to outliers, as demonstrated by the large difference between the estimates above when 6 out of 7627 observations are removed from the sample. Secondly, it would be difficult to find a simple model for this situation. An alternative estimation strategy would be separate modelling of the six large observations and then concatenation of the resulting two estimates. When using $\bar{X}_n$, such a strategy would result in just using the whole sample. For $\hat{M}$, though, this might not necessarily be the case. The problem then would be that six observations would not be enough for making a good estimate.

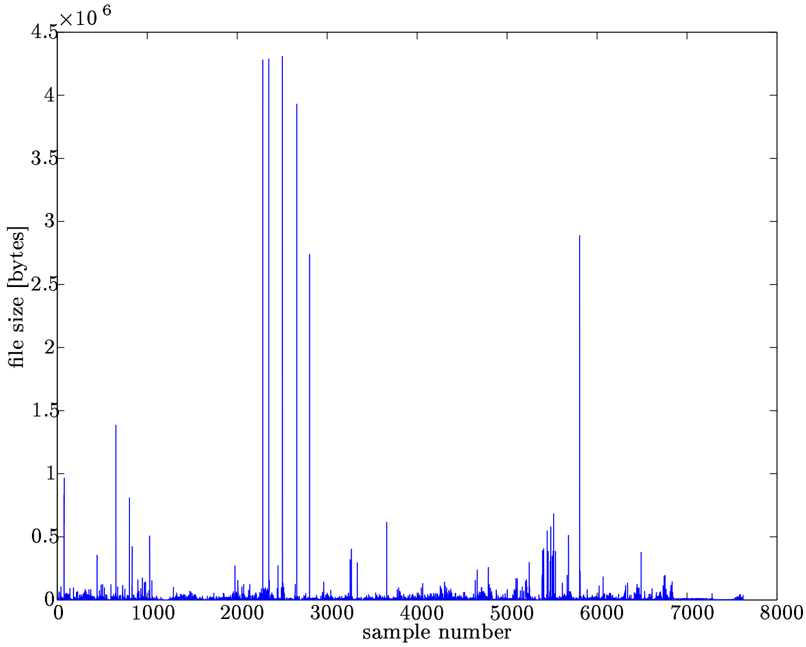

Figure 4.1: File sizes [bytes].

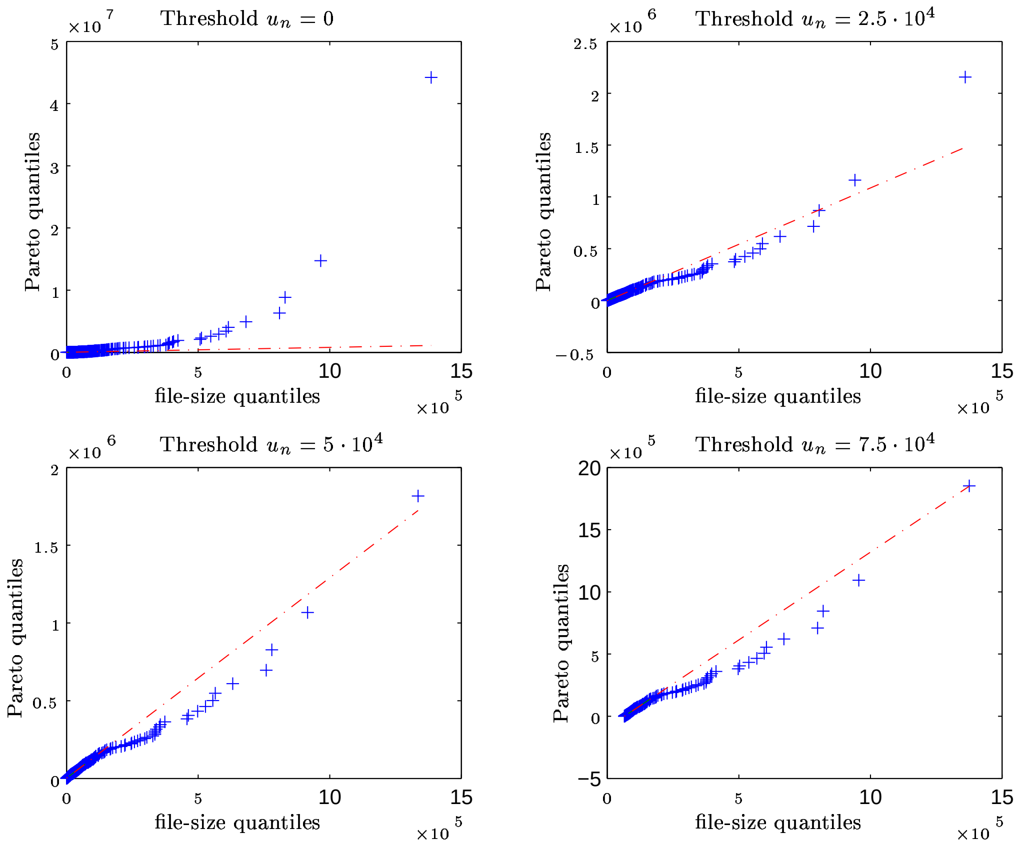

Figure 4.2: QQ-plot of data against a generalised Pareto distribution. The six largest observations have been removed.

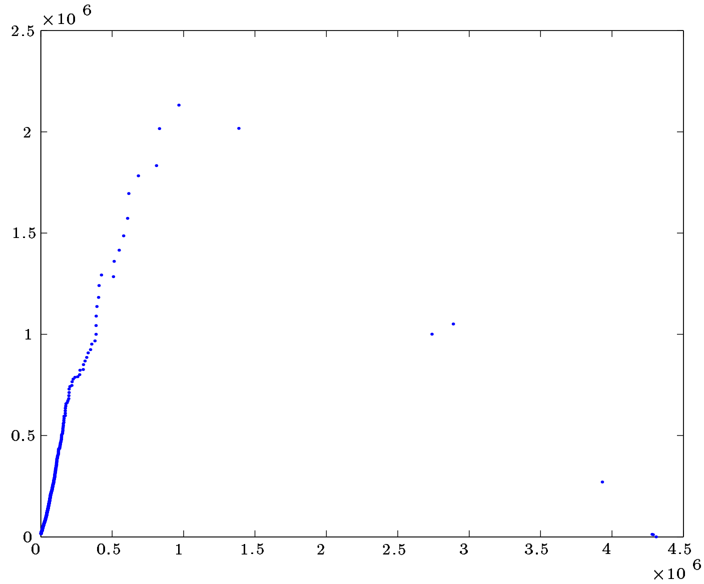

Figure 4.3: The mean-excess plot for the data suggests modelling of the file-sizes with a Pareto distribution, apart from the six largest observations.

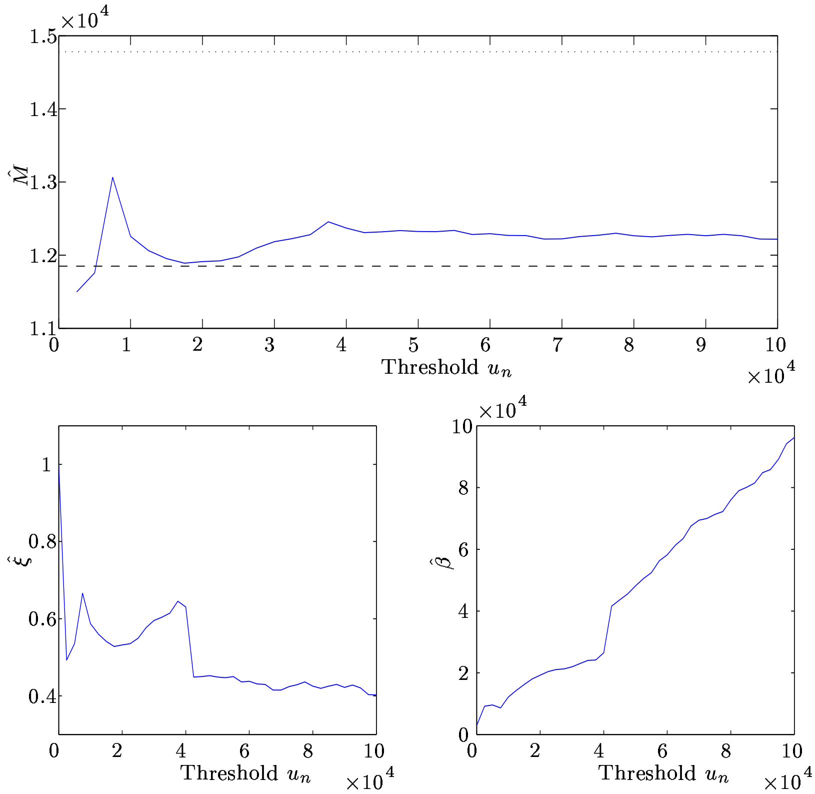

Figure 4.4: *Top:* Estimation of the mean for different values of the threshold u_n, compared to the sample mean, $\bar{X}_n$, based on all observations (dotted line) and with the six largest observations removed (dashed line). *Bottom left:* Estimation of ξ for different values of u_n. *Bottom right:* Estimation of β for different values of u_n.

4.2 A regression-type application

This section presents a simple regression-type application of the mean-estimation procedure described earlier. This procedure will, in turn, be compared to the standard methods otherwise used.

The dataset to be investigated is a portion of the publically available measurements of HTTP data gathered from the Home IP service offered by UC Berkeley to its students, faculty and staff. The dataset is described in detail on http://ita.ee.lbl.gov /html/contrib/UCB.home-IP-HTTP.html.

Due to the size of the dataset, only the measurements between Wed Nov 6 12:46:59 1996 and Sat Nov 9 20:47:01 1996, were used. And of these, only the GET requests were looked at. This left 1 577 582 observations of file down-loads to be examined.

The dataset was then divided into 10 minute intervals and the mean file-size was estimated for each of these. The division into small intervals was done so that trends and other non-stationarities would be negligible in each small interval. Then, estimates of the parameters β and ξ, as in Equation (2.5) were calculated for different thresholds u_n. Which threshold to choose was based on examination of qq- and probability-plots and on comparing the excesses over the treshold to a generalised Pareto distribution using a Kolmogorov-Smirnov test. The threshold for each interval was selected so that it passed the test with at the 5% significance level. The result is presented in Figure 4.5.

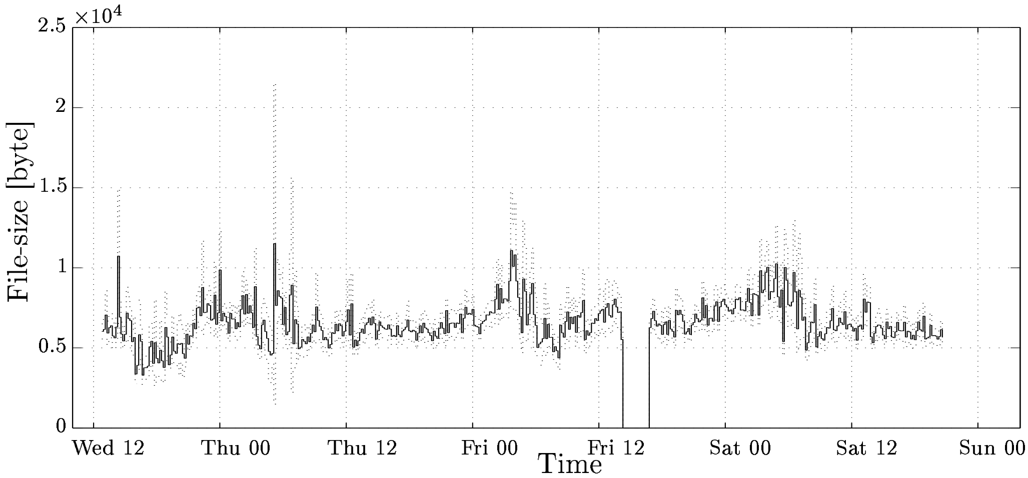

Figure 4.5: UC Berkeley mean file-sizes over 10 minute intervals. Dotted lines are 95% pointwise confidence intervals. Note that the network was down for approximately 2.5 hours on Friday afternoon. For the cases where $\hat{\xi} < 0$, the mean was estimated using the sample mean. No confidence interval is supplied for these observations.

The estimation procedure used to generate Figure 4.5 should be compared to the one using the sample mean for the same disjoint intervals. In Figure 4.6 a plot of the difference between these estimates is displayed. Note that the sample average tends to yield larger estimates for this data set.

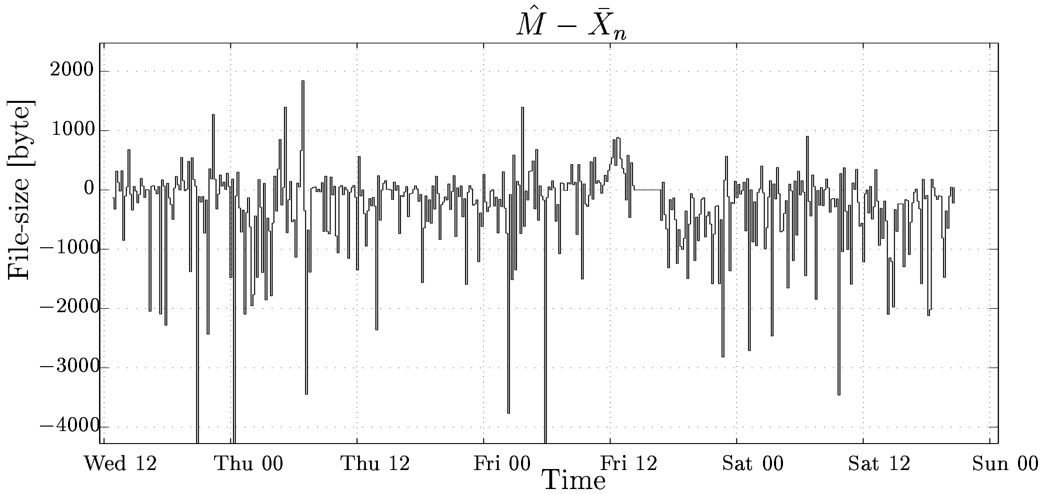

Figure 4.6: Difference between the two estimates, $\hat{M} - \bar{X}_n$. Not completely shown are two large negative observations of size $-3.5 \cdot 10^5$ and $-1 \cdot 10^5$ around Thursday midnight.

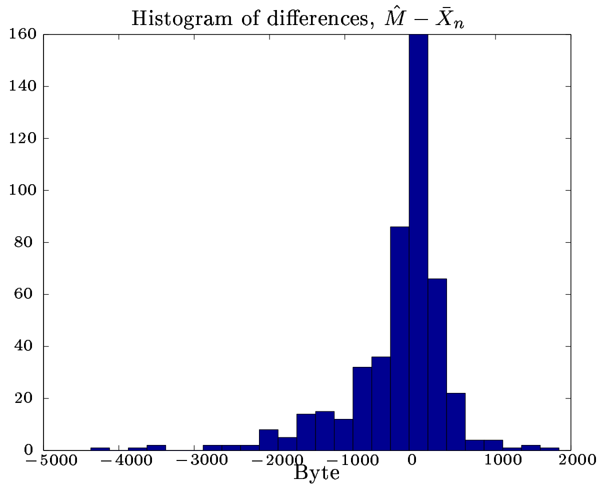

Figure 4.7: Histogram of the differences between the estimates, $\hat{M} - \bar{X}_n$. The two largest negative observations are not shown.

The number of observations used for estimating the tail-parameters ξ and β is displayed in Figure 4.9. Note that few observations were used on Wednesday afternoon, due to a drop in overall traffic. The relatively few observations, in turn, lead to greater variability in the estimate of the mean file-size for the same period.

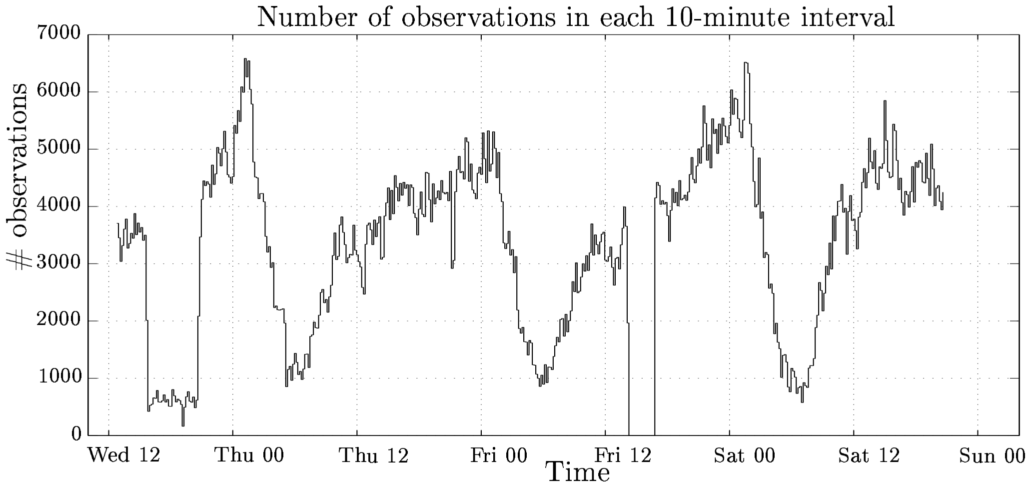

Figure 4.8: Number of observations in each 10-minute interval.

Examining the results, it appears that file-sizes increase slightly after midnight. This might indicate that users wait for periods of low traffic intensity before downloading large files, thus making the downloads faster.

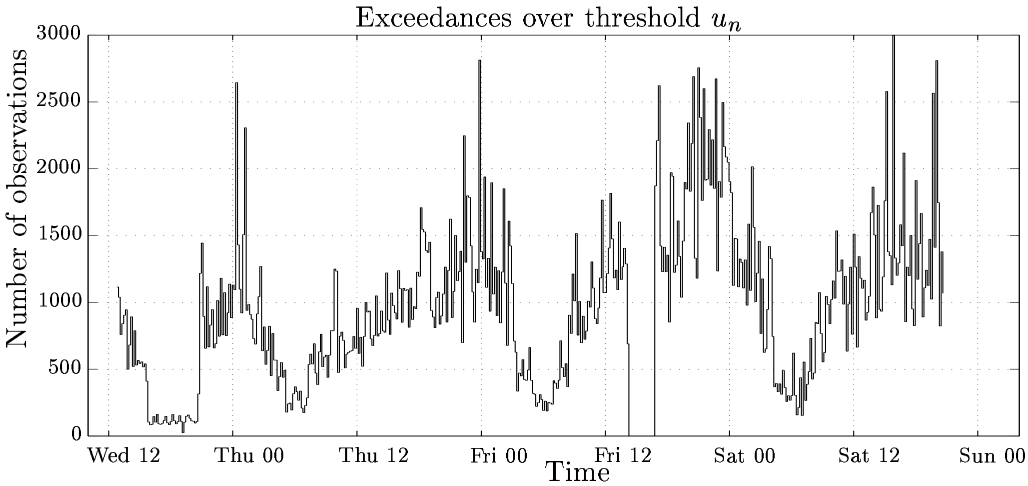

Figure 4.9: Number of observations exceeding the selected threshold. This corresponds to the number of observations that were used for estimating the parameters ξ and β.

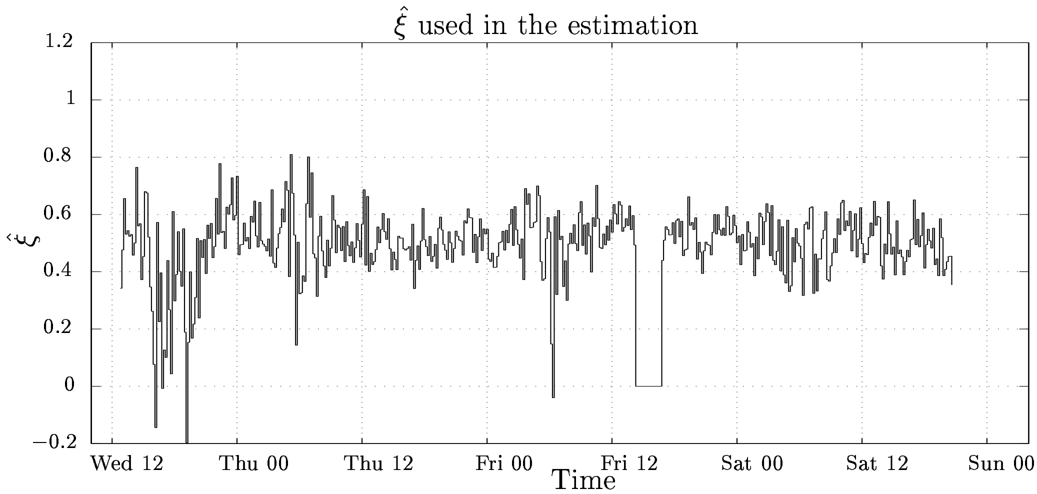

Figure 4.10: The estimated value of ξ for each of the 10-minute intervals. When $\hat{\xi} < 0$, the mean was estimated using the sample mean.

A The GPD distribution

A.1 Properties of the GPD

In order to determine the asymptotic distribution we need the information matrix for the estimate $(\hat{\beta}, \hat{\xi})$. We assume that the observations used for this estimate come from a generalised Pareto distribution with density function

$$f(x) = \frac{1}{\beta}\left(1 + \xi \frac{x}{\beta}\right)^{-1-1/\xi}, \quad \xi > 0, x \in [0, \infty)$$

Let $l(\beta, \xi) := \ln f(x)$ be the log-likelihood function and $U = (U_\beta, U_\xi)^t$ be the score function, where

$$U_\beta = \frac{1}{\beta\xi} - \frac{1}{\beta}\Big(\frac{1}{\xi} + 1\Big)\Big(1 + \xi \frac{y}{\beta}\Big)^{-1} \tag{A.1}$$

$$U_\xi = \frac{1}{\xi^2} \ln \Big(1 + \xi \frac{y}{\beta}\Big) - \frac{\xi + 1}{\xi^2}\Big(1 - \Big(1 + \xi \frac{y}{\beta}\Big)^{-1}\Big). \tag{A.2}$$

Calculation of the second derivatives results in

$$
\begin{aligned}
J_{\beta\beta} &= -\frac{\partial^2}{\partial\beta^2} l(\beta,\xi) \\
&= \frac{1}{\beta^2\xi} - \frac{1+\xi}{\beta^2\xi}\frac{1}{\left(1+\xi\frac{x}{\beta}\right)^2} && \text{(A.3)} \\
J_{\beta\xi} &= -\frac{\partial^2}{\partial\beta\partial\xi} l(\beta,\xi) \\
&= \frac{1}{\beta\xi^2} + \frac{1}{\beta\xi}\left(1+\frac{1}{\xi}\right)\frac{1}{\left(1+\xi\frac{x}{\beta}\right)^2} - \frac{1}{\beta\xi}\left(1+\frac{2}{\xi}\right)\frac{1}{\left(1+\xi\frac{x}{\beta}\right)} && \text{(A.4)} \\
J_{\xi\xi} &= -\frac{\partial^2}{\partial\xi^2} l(\beta,\xi) \\
&= \frac{2}{\xi^2}\left(1+\frac{2}{\xi}\right)\frac{1}{\left(1+\xi\frac{x}{\beta}\right)} + \frac{2}{\xi^3}\ln\left(1+\xi\frac{x}{\beta}\right) - \frac{1}{\xi^2}\left(1+\frac{1}{\xi}\right)\frac{1}{\left(1+\xi\frac{x}{\beta}\right)^2} - \\
&\quad \frac{1}{\xi^2}\left(1+\frac{3}{\xi}\right) && \text{(A.5)}
\end{aligned}
$$

We note that if X has density $f(x)$, then

$$
\begin{aligned}
\mathsf{E}\left[\left(1+\xi\frac{X}{\beta}\right)^{-r}\right] &= \frac{1}{1+r\xi}, \quad \text{if } r\xi > -1, r \in \mathbb{N} \\
\mathsf{E}\left[\ln\left(1+\xi\frac{X}{\beta}\right)\right] &= \xi
\end{aligned}
$$

This makes $\mathsf{E}[U_\beta] = \mathsf{E}[U_\xi] = 0$ and the information matrix, Q, will have components

$$
\begin{aligned}
Q_{\beta\beta} := \mathsf{E}J_{\beta\beta} &= \frac{1}{\beta^2(1+2\xi)} \\
Q_{\beta\xi} := \mathsf{E}J_{\beta\xi} &= \frac{1}{\beta(1+\xi)(1+2\xi)} \\
Q_{\xi\xi} := \mathsf{E}J_{\xi\xi} &= \frac{2}{(1+\xi)(1+2\xi)}
\end{aligned}
$$

hence

$$
Q = \begin{pmatrix} Q_{\beta\beta} & Q_{\beta\xi} \\ Q_{\beta\xi} & Q_{\xi\xi} \end{pmatrix} = \frac{1}{1+2\xi}\begin{pmatrix} \frac{1}{\beta^2} & \frac{1}{\beta(1+\xi)} \\ \frac{1}{\beta(1+\xi)} & \frac{2}{(1+\xi)} \end{pmatrix} \tag{A.6}
$$

and

$$
Q^{-1} = (1+\xi)\begin{pmatrix} 2\beta^2 & -\beta \\ -\beta & 1+\xi \end{pmatrix}
$$

B Perturbed GPD distribution

B.1 The perturbed GPD

Our model of the distribution function is

$$\bar{F}(x) = cx^{-1/\xi}(1 + x^{-\delta}L(x)),$$

where $\delta > 0$ is assumed to be known, $\xi \in (1/2, 1)$, $L(x) \in RV_0$ is a slowly varying function at infinity and c is a constant. This means that $F \in MDA(\Phi_\xi)$, where

$$\Phi_\xi(x) = \begin{cases} 0, & x < 0 \\ \exp\{-x^{-\xi}\}, & x \geq 0 \end{cases}$$

is the first extreme value distribution, see for instance Theorem $8.13.2 - 4$ in [3] or Chapter 1.2 in [14].

The distribution for the excesses over the threshold $u = u_n$ is

$$\bar{F}_u(y) = \frac{\bar{F}(u+y)}{\bar{F}(u)} = \left(1 + \frac{y}{u}\right)^{-1/\xi} \frac{1 + (u+y)^{-\delta}L(u+y)}{1 + u^{-\delta}L(u)}$$

and if $\beta = u\xi$ we note that this equals a perturbed GPD distribution.

We are interested in estimating the parameters β and ξ. Let $Y_1, Y_2, \ldots$ be the excesses over the threshold u. Use these for estimating the parameters β and ξ in the GPD by the ML method. How big an error does this lead to when the actual distribution is not GPD but rather F_u? The score function and the information matrix can both be expressed in terms of

$$\mathsf{E}\left[\ln\left(1 + \frac{Y}{u}\right)\right] \quad \text{and} \quad \mathsf{E}\left[\left(1 + \frac{Y}{u}\right)^{-r}\right], \quad r \in \{1, 2, 3 \ldots\}$$

So, if the log-likelihood function is

$$l(\beta,\xi) = \ln\Big[\frac{1}{\beta}\Big(1+\xi\frac{y}{\beta}\Big)^{-1-1/\xi}\Big],$$

this means that, for $r \in \{1,2,3,\dots\}$,

$$\begin{aligned}
\mathsf{E}\Big[\Big(1+\frac{Y}{u}\Big)^{-r}\Big] &= \frac{1}{1+r\xi} - \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\cdot r\Big(\int_1^\infty t^{-r-1-1/\xi-\delta}\frac{L(tu)}{L(u)}dt - \frac{\xi}{r\xi+1}\Big)\\
&\sim \frac{1}{1+r\xi} - \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\cdot\Big(\frac{r\xi}{\xi(r+\delta)+1} - \frac{r\xi}{r\xi+1}\Big)\\
&= \frac{1}{1+r\xi} + \frac{1}{1+r\xi}\cdot\frac{r^2\xi\delta}{(r+\delta)\xi+1}(u^{-\delta}L(u) - u^{-2\delta}L^2(u)+\dots),\\
&\qquad\qquad\text{for } |u^{-\delta}L(u)|<1\\
&\sim \frac{1}{1+r\xi} + \frac{1}{1+r\xi}\cdot\frac{r^2\xi\delta}{(r+\delta)\xi+1}u^{-\delta}L(u)\\
&\sim \frac{1}{1+r\xi}\\
\mathsf{E}\Big[\ln\Big(1+\frac{Y}{u}\Big)\Big] &= \xi + \int_1^\infty t^{-1-1/\xi}\frac{(tu)^{-\delta}L(tu) - u^{-\delta}L(u)}{1+u^{-\delta}L(u)}dt\\
&\sim \xi - \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\cdot\frac{\xi\delta}{1+\xi\delta}\\
&= \xi - \xi\cdot\frac{\delta}{1+\xi\delta}(u^{-\delta}L(u) - u^{-2\delta}L^2(u)+\dots),\quad \text{for } |u^{-\delta}L(u)|<1\\
&\sim \xi - \xi\cdot\frac{\delta}{1+\xi\delta}u^{-\delta}L(u)\\
&\sim \xi
\end{aligned}$$

where $|u^{-\delta}L(u)| < 1$ is needed for the series expansion to work. Making use of Equations A.1 and A.2 in Appendix A and calculating the expectation of U_β and U_ξ using, not the df for GPD, but rather F_u, results in

$$\begin{aligned}
\mathsf{E}[U_\beta] &\sim -\frac{\delta\xi}{\beta(1+\xi(1+\delta))}\cdot\frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\\
&= -\frac{\delta\xi}{\beta(1+\xi(1+\delta))}\cdot(u^{-\delta}L(u) - u^{-2\delta}L^2(u)+\dots)\\
\mathsf{E}[U_\xi] &\sim \frac{\delta^2\xi(\xi-1)-\delta\xi}{\xi(1+\delta\xi)(1+\xi(1+\delta))}\cdot\frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\\
&= \frac{\delta^2\xi(\xi-1)-\delta\xi}{\xi(1+\delta\xi)(1+\xi(1+\delta))}\cdot(u^{-\delta}L(u) - u^{-2\delta}L^2(u)+\dots),
\end{aligned}$$

for u sufficiently large so that $|u^{-\delta}L(u)| < 1$. Analogously, using Equations A.3 – A.5 in Appendix A,

$$
\begin{aligned}
Q_{\beta\beta} = \mathsf{E}[J_{\beta\beta}] &\sim \frac{1}{\beta^2(1+2\xi)} - \frac{2\delta\xi(1+\xi)}{\beta^2(1+2\xi)(1+(2+\delta)\xi)} \cdot \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)} \\
&\sim \frac{1}{\beta^2(1+2\xi)} - \frac{1}{\beta^2(1+2\xi)} \cdot \frac{2\delta\xi(1+\xi)}{(1+(2+\delta)\xi)} u^{-\delta}L(u) \\
Q_{\beta\xi} = \mathsf{E}[J_{\beta\xi}] &\sim \frac{1}{\beta(1+\xi)(1+2\xi)} - \frac{\delta\xi(3+(6+\delta)\xi+2\xi^2)}{\beta(1+\xi)(1+2\xi)(1+\xi(1+\delta))(1+\xi(2+\delta))} \\
&\quad \cdot \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)} \\
&\sim \frac{1}{\beta(1+\xi)(1+2\xi)} \Big(1 - \frac{\delta\xi(3+(6+\delta)\xi+2\xi^2)}{(1+\xi(1+\delta))(1+\xi(2+\delta))} u^{-\delta}L(u) \Big) \\
Q_{\xi\xi} = \mathsf{E}[J_{\xi\xi}] &\sim \frac{2}{(1+\xi)(1+2\xi)} + \frac{2\delta}{\xi^2} \Big(-\frac{1}{1+\delta\xi} + \frac{\xi(2+\xi)}{(1+\xi)(1+\xi(1+\delta))} \\
&\quad - \frac{\xi(1+\xi)}{(1+2\xi)(1+(2+\delta)\xi)} \Big) \cdot \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)} \\
&\sim \frac{2}{(1+\xi)(1+2\xi)} + \frac{2}{(1+\xi)(1+2\xi)} \cdot \\
&\frac{\delta[\delta(3+\delta)\xi^5 + (\delta^2+4\delta-1)\xi^4 - (2\delta^2+4\delta+3)\xi^3 - (\delta^2+7\delta+7)\xi^2 - (5+2\delta)\xi - 1]}{(1+\delta\xi)(1+\xi(1+\delta))(1+\xi(2+\delta))} \cdot \\
&\quad u^{-\delta}L(u),
\end{aligned}
$$

resulting in an information matrix

$$
Q = \begin{pmatrix} Q_{\beta\beta} & Q_{\beta\xi} \\ Q_{\beta\xi} & Q_{\xi\xi} \end{pmatrix}, \tag{B.1}
$$

where u is assumed to be sufficiently large so that $|u^{-\delta}L(u)| < 1$, as above.

B.1.1 Q^{-1}

The aim is to estimate Q^{-1} in such a way that it contains only constants and terms of order $u^{-\delta}L(u)$. To do this, observe that we can write $Q = Q_{GPD} + Q_{Pert}$, where Q_{GPD} is the information matrix we get when the underlying distribution is exactly Pareto (c.f. Equation A.6 in Appendix A), and Q_{Pert} is the contribution from the perturbation, i.e. the contribution arising from not having an exact Pareto distribution.

Now approximate Q^{-1} in the following way:

$$
\begin{aligned}
Q^{-1} &= (Q_{GPD} + Q_{Pert})^{-1} = (\mathbf{1} + Q_{GPD}^{-1}Q_{Pert})^{-1}Q_{GPD}^{-1} = \\
&= \sum_{k=0}^{\infty} (-Q_{GPD}^{-1}Q_{Pert})^k Q_{GPD}^{-1} = Q_{GPD}^{-1} - Q_{GPD}^{-1}Q_{Pert}Q_{GPD}^{-1} + R \approx \\
&\approx Q_{GPD}^{-1} - Q_{GPD}^{-1}Q_{Pert}Q_{GPD}^{-1},
\end{aligned}
$$

where R is the remainder of the sum. The advantage to this is that all elements in R will be functions of $(u^{-\delta}L(u))^k$, for $k > 1$ and hence should be negligible. Another advantage is that there is no need to explicitly calculate Q_{Pert}^{-1}, which would have been a messy affair.

Does $\sum(-Q_{GPD}^{-1}Q_{Pert})^k$ converge? It will if $\|Q_{GPD}^{-1}Q_{Pert}\|_2 < 1$. This is shown the following way:

$$\begin{aligned}
\|Q_{GPD}^{-1}Q_{Pert}\|_2^2 &\leq \left|\frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\right|^2 \cdot \|Q_{GPD}^{-1}\|_2^2 \cdot \|Q'_{Pert}\|_2^2 \\
&= \left|\frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\right|^2 \cdot \lambda_{\max}((Q_{GPD}^{-1})^t Q_{GPD}^{-1}) \cdot \lambda_{\max}((Q'_{Pert})^t Q'_{Pert}),
\end{aligned}$$

where $(u^{-\delta}L(u)/(1+u^{-\delta}L(u)))Q'_{Pert} = Q_{Pert}$ and $\lambda_{\max}(A)$ is the largest eigenvalue of A. Finally, find bounds for the eigenvalues by using

Theorem B.1 (Gershgorin's Theorem)
Let B be an arbitrary matrix. Then the eigenvalues λ of B are located in the union of the n disks defined by

$$|\lambda - b_{ii}| \leq \sum_{j\neq i} |b_{ij}|, \quad \text{for } i = 1, \dots, n.$$

Proof: This is Theorem 2.9 in [5]. □

The matrix we are interested in finding the eigenvalues for is

$$M = (Q'_{Pert})^t Q'_{Pert}$$

and since Q'_{Pert} is a symmetric matrix, we can write

$$Q'_{Pert} = PDP^t \quad \Rightarrow \quad M = PD^2P^t,$$

where P is the matrix with columns being the eigenvectors of Q'_{Pert} and $D = \mathrm{diag}\{\lambda_1, \lambda_2\}$ is the matrix with the corresponding eigenvalues in the diagonal. Hence, the eigenvalues of M are all positive.

The conclusion is that the eigenvalues of the perturbation matrix have to be smaller than both

$$100\frac{\delta^2}{\beta^4+\beta^2} + 1000\frac{\delta^2+\delta^5}{\beta} \quad \text{and} \quad 320\delta^2 + \frac{2500}{(\delta+\delta^2)(\beta^2+\beta^3)},$$

making

$$\lambda_{\max}((Q'_{Pert})^t Q'_{Pert}) \leq \max\left(100\frac{\delta^2}{\beta^4+\beta^2} + 1000\frac{\delta^2+\delta^5}{\beta}, 320\delta^2 + \frac{2500}{(\delta+\delta^2)(\beta^2+\beta^3)}\right).$$

Using the same method on the remaining eigenvalue results in

$$\lambda_{\max}((Q_{GPD}^{-1})^t Q_{GPD}^{-1}) \leq 16 + 4\beta^2 + 16\beta^4.$$

This all illustrates that we will have large values of the bound for both large and small values of β and δ, assuming $\xi \in (0.5, 1)$. Still, we have an (albeit coarse) limit for the eigenvalues demonstrating that it is possible to get convergence in the approximation of the inverse by selecting a large enough value of u.

To summarise the above,

$$\|Q_{GPD}^{-1} Q_{Pert}\|_2^2 \leq \left|\frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)}\right|^2 \cdot (16 + 4\beta^2 + 16\beta^4) \cdot \max\left(100\frac{\delta^2}{\beta^4+\beta^2} + 1000\frac{\delta^2+\delta^5}{\beta}, 320\delta^2 + \frac{2500}{(\delta+\delta^2)(\beta^2+\beta^3)}\right).$$

So we arrive at an approximate expression for Q^{-1}:

Lemma B.1 (Approximation of Q^{-1})

$$\begin{aligned} Q^{-1} &= Q_{GPD}^{-1} - Q_{GPD}^{-1} Q_{Pert} Q_{GPD}^{-1} + O\left(\|Q_{GPD}^{-1} Q_{Pert}\|_2^2\right) \\ &\approx Q_{GPD}^{-1} - Q_{GPD}^{-1} Q_{Pert} Q_{GPD}^{-1}, \end{aligned}$$

for large u, where

$$Q_{GPD}^{-1} = (1+\xi)\begin{pmatrix} 2\beta^2 & -\beta \\ -\beta & 1+\xi \end{pmatrix}$$

and

$$Q_{GPD}^{-1} Q_{Pert} Q_{GPD}^{-1} = \frac{\delta(1+\xi)}{\xi^2(1+\delta\xi)(1+(1+\delta)\xi)(1+(2+\delta)\xi)} \begin{pmatrix} (1,1) & (1,2) \\ (2,1) & (2,2) \end{pmatrix} \frac{u^{-\delta}L(u)}{1+u^{-\delta}L(u)},$$

with

$$\begin{aligned} (1,1) &= -2\beta^2(1 + \xi(3 + 2\delta + \xi + \delta(3+\delta)\xi - (1+2\delta)\xi^2 - (\delta^2-3)\xi^3 + \\ &\quad (1+\delta)(2+3\delta)\xi^4 + 2\delta(1+\delta)\xi^5)) \\ (1,2) = (2,1) &= \beta(2 + 4(2+\delta)\xi + 2(1+\delta)(4+\delta)\xi^2 + (2\delta(1+\delta) - 1)\xi^3 - \\ &\quad 2(1+\delta(4+\delta))\xi^4 + (\delta-4)\delta\xi^5 + 2\delta^2\xi^6) \\ (2,2) &= 2(1+\xi)(-1 + \xi(-4(1+\xi) + \delta(-2 + \xi(\delta((\xi-1)\xi - 1) \\ &\quad + (\xi-1)(5 + 2\xi(3+\xi)))))) \end{aligned}$$

Bibliography

[1] R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors. *A Practical Guide to Heavy Tails.* Birkhäuser, 1998.

[2] P. Billingsley. *Probability and Measure.* Wiley, 3rd edition, 1995.

[3] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Number 27 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1987.

[4] M. E. Crovella, M. S. Taqqu, and A. Bestavros. Heavy-Tailed Probability Distributions in the World Wide Web. In R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails*, pages 3–25. Birkhäuser, 1998.

[5] J. W. Demmel. *Applied Numerical Linear Algebra.* SIAM, 1997.

[6] R Durrett. *Probability: Theory and Examples, Second Edition.* Duxbury Press, 1996.

[7] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events.* Springer, 1997.

[8] C. A. Guerin, H. Nyberg, O. Perrin, S. Resnick, H. Rootzén, and C. Starica. Empirical Testing of the Infinite Source Poisson Data Traffic Model. Preprint, Chalmers University of Technology, Sweden, no 2000:4.

[9] F. P. Kelly, S. Zachary, and I. Ziedins, editors. *Stochastic Networks, Theory and Applications*, volume 4 of *Royal Statistical Society Lecture Note Series.* Oxford University Press, 1996.

[10] J. B. Levy and M. S. Taqqu. Renewal Reward Processes with Heavy-Tailed Inter-Renewal Times and Heavy-Tailed Rewards. *Bernoulli*, 6(1):23–44, 2000.

[11] J. K. Lindsey. *Parametric Statistical Inference.* Oxford University Press, 1996.

[12] I. Norros. On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks. *IEEE Journal on Selected Areas in Communications*, 13(6):953–962, August 1995.

[13] V. Pipiras and M. S. Taqqu. The Limit of a Renewal Reward Process with Heavy-Tailed Rewards is not a Linear Fractional Stable Motion. *Bernoulli*, 6(4):607–614, 2000.

[14] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, 1987.

[15] J. P. Romano and M. Wolf. Subsampling inference for the mean in the heavy-tailed case. *Metrika*, 50:55–69, 1999.

[16] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman & Hall, 1994.

[17] E. Seneta. *Regularly Varying Functions.* Number 508 in Lecture Notes in Mathematics. Springer-Verlag, 1976.

[18] R.L. Smith. Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika*, 72:67–90, 1985.

[19] R.L. Smith. Estimating Tails of Probability Distributions. *The Annals of Statistics*, 15(3):1174–1207, 1987.

[20] N. Tajvidi. *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions.* PhD thesis, Chalmers University of Technology, 1996.

[21] A.W. van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistical Mathematics. Cambridge University Press, 1998.