THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Contributions to Quantile Estimation

Erik Brodin

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
Göteborg, Sweden 2005

Contributions to Quantile Estimation
Erik Brodin

NO 2005:1
ISSN 0347-2809

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Printed in Göteborg, Sweden 2005

Contributions to Quantile Estimation

Erik Brodin

Department of Mathematical Statistics
Chalmers University of Technology
Göteborg University

Abstract

In this thesis we develop new methodology for quantile estimation.

For quantiles within the range of the sample we use non-parametric estimation, based on linear combinations of order statistics, so called L-estimators. In our first article, we improve on the Harrell-Davis estimator by smoothing techniques. We also prove central limit theorems for the Harrell-Davis estimator and our improvements of it. In our second article, we use exact bootstrap to construct an optimal L-estimator in the mean square error sense.

For quantiles out of the sample we use parametric estimation, based on second order regular variation techniques. The purpose of this is to lower bias resulting from poor speed of convergence, by means of incorporating that speed into the model. In our third article, we use second order regular variation together with cross validation. In our fourth article, we use perturbed Generalized Pareto distributions to model second order regular variation.

Keywords: Bootstrap; Cross validation; Extreme value theory; Generalized Pareto distribution; Kernel estimator; L-estimator; Non-parametric estimation; Order statistic; Perturbed Generlized Pareto distribution; Quantile estimation; Regular variation; Second order regular variation.

Acknowledgements

I am very gratefull to my supervisor Professor Patrik Albin who has an energy which is beyond my understanding. The encouragement and support of Patrik as well as his comments about my work have been invaluable.

I would also like to thank my co-supervisor Professor Holger Rootzén who has been an inspiration in extreme value theory.

My colleagues at the Department of Mathematical Sciences have been a great company and formed a stimulating environment. Among a few, Professors Rossitza Dodunekova and Dragi Anevski, the lunch table, my room mate Carl and the palbin gang. My colleagues and friends Erik Broman and Martin Adiels are worth a special thanks for many non work-related and amusing discussions.

Finally, I would like to thank my family for inspiration and support. Especially my fiancée Hanna.

Göteborg, January 27, 2005
Erik Brodin

This thesis consists of the following papers:

Paper I: *Smoothing the Harrell-Davis Quantile Estimator,* joint work with J.M.P Albin. Submitted to Journal of Scandinavian Statistics.

Paper II: *On Quantile Estimation by Bootstrap.* Submitted to Computational Statistics & Data Analysis.

Paper III: *Estimating Extreme Quantiles using Cross Validation and Second Order Regular Variation.* Submitted to Journal of the Royal Statistical Society: Series B (Statistical Methodology).

Paper IV: *A Note on Quantile Estimation using Perturbed Generalized Pareto Distributions.* Submitted to Journal of Statistical Computation and Simulation.

PAPER I

Smoothing the Harrell-Davis Quantile Estimator

J.M.P Albin† and Erik Brodin

Chalmers University of Technology

27th January 2005

Abstract

Quantile estimation can made more efficiently by using linear combinations of order statistics, so called L-estimators, instead of a single sample quantile. One estimator of this type is the Harrell-Davis estimator, which can be viewed as a kernel smoothing of the empirical quantile function. We consider modifications of this estimator by smoothing over more complicated estimators than the empirical quantile function. We show that our proposed new estimators, as well as the original Harrell-Davis estimator obey the same central limit theorem as does the empirical quantile.

Keywords and phrases: Kernels, L-estimators, Non-parametric estimation, Order statistics, Quantile estimation

1 Introduction

Non-parametric estimatiors of population quantiles are of fundamental importance in statistics: Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, with probability distribution function F. Let $X_{(1)} \leq \ldots \leq X_{(n)}$ be the corresponding ordered sample. We define the quantile function Q as the left inverse of F, given by

$$Q(p) = \inf\{x : F(x) \geq p\} \quad \text{for } 0 < p < 1.$$

A basic estimator of $Q(p)$, the p:th quantile, is the p:th sample quantile, given by $\hat{Q}_n(p) = X_{(\lfloor np \rfloor + 1)}$, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$: One has, under mild regularity conditions, and with obvious notation,

$$\sqrt{n}\,(\hat{Q}_n(p) - Q(p)) \xrightarrow{d} \mathrm{N}\left(0, \frac{p(1-p)}{f(Q(p))^2}\right) \quad \text{as } n \to \infty \text{ for } 0 < p < 1.$$

†Research supported by The Swedish Research Council contract no. 621-2003-5214.

where f is the density function of F: See Serfling (1980), p. 77.

However, it turns out that one can improve on the estimator $\hat{Q}_n(p)$ of $Q(p)$ by averaging over the order statistics, using suitable weights w_i:

$$L_n = \sum_{i=1}^{n} w_i X_{(i)} \quad \text{where} \quad \sum_{i=1}^{n} w_i = 1.$$

These estimators are commonly called L-estimators. Notice that $\hat{Q}_n(p)$ is an L-estimator with $w_{\lfloor np \rfloor+1} = 1$ and $w_i = 0$ for $i \neq \lfloor np \rfloor+1$. Obviously, a problem occurs with choosing the weight function w_i, because it will be a trade off between bias and variance: The more order statistics that are incoperated in the L-estimator, the smaller will be the variance, but the larger will be the bias, and vice versa.

A popular class of L-estimators are kernel quantile estimators, given by

$$KQ_n(p) = \sum_{i=1}^{n} [\int_{(i-1)/n}^{i/n} K_h(t-p)dt] X_{(i)}. \tag{1}$$

Here K is a density function symmetric about zero, while $h = h(n) \to 0$ as $n \to \infty$, and $K_h(\cdot) = K(\cdot/h)/h$. Notice the delicate task of choosing h!

Kernel quantile estimators have been studied intensely, and we refer to Sheather and Marron (1990) for more information and for additional references.

Notice that the weights in front of $X_{(i)}$ in the sum (1) do not add up to 1 in general. However, it has been shown, that under some regularity conditions, an L-estimator with weights standardized to add up to 1 (not only of kernel type), has the same asymptotic distribution as the corresponding unstandardized estimator, see Sheather and Marron (1990), Theorem 2, and Lehmann (1983), Corollary 5.1: And so the weights in the sum (1) do not really have to add up to 1!

Also, it has been shown that if the kernel is well selected, then the estimator will have the same asymptotic behaviour as has the sample quantile, see Yang (1985).

Another L-estimator, with a different motivation, is the Harrell-Davis estimator, proposed by Harrell and Davis (1982): Notice that the expected value of the k:th order statistic $X_{(k)}$, is given by

$$\mathbf{E}\{X_{(k)}\} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \int_0^1 Q(y) y^{k-1} (1-y)^{n-k} dy.$$

Observing that $\mathbf{E}\{X_{((n+1)p)}\} \to Q(p)$ as $n \to \infty$ for $p \in (0,1)$, see David (1981) p. 80, Harrell and Davis proposed the following estimator of $Q(p)$:

$$\text{HD}_n(p) = \frac{\Gamma(n+1)}{\Gamma((n+1)p)\Gamma((n+1)(1-p))} \int_0^1 \hat{Q}_n(y) y^{(n+1)p-1} (1-y)^{(n+1)(1-p)-1} dy. \tag{2}$$

This estimator coincides with the exact bootstrap estimate of $\mathbf{E}\{X_{((n+1)p)}\}$, as was shown by Hutson and Ernst (2000), p. 91.

Making use of the fact that $\hat{Q}_n$ is the inverse of the empirical distribution function,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}},$$

the HD estimator can be expressed as an L-estimator:

$$\mathrm{HD}_n(p) = \sum_{i=1}^{n} w_{n,i}(p) X_{(i)}, \tag{3}$$

where (with B denoting the beta function)

$$w_{n,i}(p) = \frac{1}{B\left(p(n+1), (1-p)(n+1)\right)} \int_{(i-1)/n}^{i/n} y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1} dy.$$

Harrell and Davis claimed a central limit theorem for their estimator, relying on a result of David (1981) p. 273. But their argument seems to be incomplete, as the L-estimator cannot be written on the form $L_n = n^{-1}\sum_{i=1}^{n} J(\frac{i}{n})X_{(i)}$, for some function J, as is required by the result of David. Indeed, Yoshizawa, Sen and Davis (1985) noticed that the claimed asymptotic normality was incorrect, and proved asymptotic normality for $p = 0.5$, the median.

It was shown by Zelterman (1990), p. 343, building in part on results and arguments from Sheather and Marron (1990) and Yang (1985), that the HD estimator yields to the same central limit theorem as do $\hat{Q}_n$, by means of identifying that the weights $w_{n,i}$ in (3) behave asymptotically as a normal kernel. However, we have not been able to follow Zelterman's arguments: For example, the estimate of the variance, on the lower part of Zelterman (1990), p. 343, seems to be erroneous, as one cannot lift out γ_i from the sum, as is done there. Also, the argument for the mentioned central limit theorem, that proceeds the variance estimate, is sketchy indeed, and we cannot follow Zelterman here, based on the arguments he provides.

We have been in contact with Professor Zelterman, and he generously does no dispute the possible unclearities in his proof that we have suggested.

2 Modified HD Estimators

We propose two modifications of the HD estimator, based on using more complicated estimates of Q that $\hat{Q}_n$. These new estimators can be viewed as "double smoothings" of the simple sample quantile estimator.

Sheather and Marron (1990), p. 412, argue that the HD estimator have weights that are too concentrated around p, to be a really optimal bandwidth, as it is asymptotically a kernel estimator: A "double smoothing" does lead to a larger bandwidth!

First modification We replace $\hat{Q}_n$ with the HD estimator on the right-hand side of (2). The resulting quantile estimator can be expressed as an L-estimator:

$$\mathrm{HD}_n^{\mathrm{HD}}(p) = \sum_{i=1}^{n} w_{n,i}^{\mathrm{HD}}(p) X_{(i)},$$

where

$$w_{n,i}^{\mathrm{HD}}(p) = \frac{1}{B\left(p(n+1),(1-p)(n+1)\right)} \int_0^1 w_{n,i}(y) y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1} dy.$$

Second modification We replace $\hat{Q}_n$ with the kernel estimator

$$\int_0^1 K_h(t-p)\hat{Q}_n(t)dt$$

on the right-hand side of (2). The resulting quantile estimator is the L-estimator

$$\mathrm{HD}_n^{\mathrm{Kernel}}(p) = \sum_{i=1}^{n} w_{n,i}^{\mathrm{Kernel}}(p) X_{(i)},$$

where

$$w_{n,i}^{\mathrm{Kernel}}(p)$$
$$= \frac{1}{B\left(p(n+1),(1-p)(n+1)\right)} \int_{(i-1)/n}^{i/n} \left(\int_0^1 K_h(x-y) y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1} dy \right) dx.$$

A problem with this estimator is the introduction of the bandwidth parameter h.

3 Asymptotic Behaviour of Modified HD Estimators

Of course, we have [see e.g., van der Vaart and Weller (1996), Theorem 3.8.1]

$$\sqrt{n}(F_n - F) \xrightarrow{d} \bar{W} \circ F \quad \text{for } 0 < p < 1, \tag{4}$$

in the sense of weak convergence in the space $D(\bar{\mathbb{R}})$ of càdlàg functions equipped with the Skorohod J_1 topology, where $\bar{W}$ denotes a standard Brownian bridge.

For an absolutely continuous distribution function F, with a strictly positive density function f, we can follow van der Vaart and Weller (1996), Example 3.9.21, to get

$$\sqrt{n}\,(\hat{Q}_n(p) - Q(p)) \xrightarrow{d} \frac{\bar{W}(F(Q(p)))}{f(Q(p))} = \frac{\bar{W}(p)}{f(Q(p))} \stackrel{d}{=} \mathrm{N}\left(0, \frac{p(1-p)}{f(Q(p))^2}\right) \quad \text{as } n \to \infty: \tag{5}$$

Since Q is not bounded in general, this convergence only is pointwise for $p \in (0,1)$. But for F with compact support, van der Vaart and Weller (1996), Example 3.9.24, gives

$$\sqrt{n}\,(\hat{Q}_n - Q) \xrightarrow{d} \frac{\bar{W}(F(Q))}{f(Q)} \stackrel{d}{=} \frac{\bar{W}}{f(Q)} \quad \text{as } n \to \infty, \tag{6}$$

in the sense of weak convergence on the space of bounded functions on $l^\infty(0,1)$ equipped with the topology of uniform convergence.

To prove a version of (5) for the modified HD estimators, we use the following lemma:

Lemma 3.1 (MASON (1982), THEOREM 3). *Let F be an absolutely continuous distribution function with a strictly positive and continuous probability density function f, such that*

$$\int_{\mathbb{R}} |x|^\alpha f(x)dx < \infty \quad \textit{for some } \alpha > 0. \tag{7}$$

We have

$$\mathbf{P}\left\{\limsup_{n\to\infty} \sup_{u\in(0,1)} u^{1/\alpha}(1-u)^{1/\alpha}|\hat{Q}_n(u) - Q(u)| = 0\right\} = 1.$$

We first give a new proof of the central limit theorem for the HD estimator. As have been mentioned, we do not trust the proofs in the literature. Further, our proofs of central limit theorems for the modified HD estimators build on the proof for the HD estimator, so that these arguments have to be supplied anyway.

Proposition 3.2 (HARRELL AND DAVIS (1982), ZELTERMAN (1990), P. 343). *Let F be an absolutely continuous distribution function with a strictly positive continuous probability density function f, such that (7) holds. The HD estimator satisfies the same central limit theorem as does $\hat{Q}$:*

$$\sqrt{n}\;(HD_n(p) - Q(p)) \xrightarrow{d} N\!\left(0, \frac{p(1-p)}{f(Q(p))^2}\right) \quad \textit{as } n \to \infty \textit{ for } p \in (0,1).$$

Proof. For measurable $f \in l^\infty(0,1)$, define, with obvious notation,

$$\begin{aligned} T_{n,p}(f) &= \frac{1}{B\,(p(n+1),(1-p)(n+1))} \int_0^1 f(y) y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1}dy \\ &= \mathbf{E}_{Y_n}\left\{f(Y_n)\right\}, \end{aligned}$$

where Y_n is a beta distributed random variable with parameters $(n+1)p$ and $(n+1)(1-p)$. By the well-known feature of beta distributions, being the empirical distribution function of samples from the uniform distribution, it is immediate from (5), that

$$\sqrt{n}\,(Y_n - p) \xrightarrow{d} \mathrm{N}\,(0, p(1-p)) \quad \text{as } n \to \infty. \tag{8}$$

We have, with obvious notation,

$$\begin{aligned}\sqrt{n}\,(\mathrm{HD}_n(p) - Q(p)) &= \sqrt{n}\,\Big(T_{n,p}(\hat{Q}_n(p)) - Q(p)\Big)\\ &= \sqrt{n}\,\mathbf{E}_{Y_n}\{\hat{Q}_n(Y_n) - Q(Y_n)\} + \sqrt{n}\,\mathbf{E}_{Y_n}\{Q(Y_n) - Q(p)\}\,.\end{aligned} \tag{9}$$

Here, for the the second term on the right hand, we have, by Taylor expansion, together with (8) and that fact that $\mathbf{E}\{Y_n\} = p$, and with obvious notation,

$$\begin{aligned}&\sqrt{n}\,\mathbf{E}_{Y_n}\left\{(Q(Y_n) - Q(p))\,I_{|Y_n - p|\leq\varepsilon}\right\}\\ &= \sqrt{n}\,\mathbf{E}_{Y_n}\left\{Q'(p)(Y_n - p)I_{|Y_n-p|\leq\varepsilon} + \int_p^{Y_n}\big(Q'(q) - Q'(p)\big)\,dq\,I_{|Y_n-p|\leq\varepsilon}\right\}\\ &= \sqrt{n}\,\mathbf{E}_{Y_n}\left\{-\frac{Y_n - p}{f(Q(p))}I_{|Y_n-p|>\varepsilon} + O\Bigg(|Y_n - p|\sup_{q\in[p-\varepsilon,p+\varepsilon]}\left|\frac{1}{f(Q(q))} - \frac{1}{f(Q(p))}\right|\Bigg)\right\}\\ &\xrightarrow{d} 0 \quad \text{as } n \to \infty \text{ and } \varepsilon \downarrow 0, \text{ in that order:}\end{aligned} \tag{10}$$

Here we also relied on the continuity of f, together with the fact that

$$\sqrt{n}\,\mathbf{E}_{Y_n}\left\{\frac{Y_n - p}{f(Q(p))}I_{|Y_n-p|>\varepsilon}\right\} \leq \sqrt{\mathbf{E}_{Y_n}\left\{\left(\frac{\sqrt{n}\,(Y_n - p)}{f(Q(p))}\right)^2\right\}\mathbf{P}_{Y_n}\{|Y_n - p| > \varepsilon\}} \xrightarrow{d} 0 \tag{11}$$

as $n \to \infty$, by the Cauchy-Schwarz inequality. Moreover, using (11) again, we get

$$\begin{aligned}&\sqrt{n}\,\mathbf{E}_{Y_n}\left\{(Q(Y_n) - Q(p))\,I_{Y_n\in[\varepsilon,p-\varepsilon)\cup(p+\varepsilon,1-\varepsilon]}\right\}\\ &\leq (Q(1-\varepsilon) - Q(\varepsilon))\sqrt{n}\,\mathbf{P}_{Y_n}\left\{Y_n \in [\varepsilon, p-\varepsilon) \cup (p+\varepsilon, 1-\varepsilon]\right\}\\ &\leq (Q(1-\varepsilon) - Q(\varepsilon))\,\mathbf{E}_{Y_n}\left\{\frac{\sqrt{n}\,|Y_n - p|}{\varepsilon}I_{|Y_n-p|>\varepsilon}\right\}\\ &\leq (Q(1-\varepsilon) - Q(\varepsilon))\sqrt{\mathbf{E}_{Y_n}\left\{\left(\frac{\sqrt{n}\,(Y_n - p)}{\varepsilon}\right)^2\right\}\mathbf{P}_{Y_n}\{|Y_n - p| > \varepsilon\}} \xrightarrow{d} 0 \quad \text{as } n \to \infty.\end{aligned} \tag{12}$$

Notice that, in particular,

$$\sqrt{n}\mathbf{P}_{Y_n}\{|Y_n - p| > \varepsilon\} \leq \sqrt{\mathbf{E}_{Y_n}\left\{\left(\frac{\sqrt{n}\,(Y_n - p)}{\varepsilon}\right)^2\right\}\mathbf{P}_{Y_n}\{|Y_n - p| > \varepsilon\}} \xrightarrow{d} 0 \quad \text{as } n \to \infty. \tag{13}$$

Finally, by (7) we have $F(-x) \leq C(-x)^{-\alpha}$ and $1 - F(x) \leq Cx^{-\alpha}$ for $x > 0$, for some constant $C > 0$, which readily gives

$$-\Big(\frac{C}{y}\Big)^{1/\alpha} \leq F^{-1}(y) \leq \Big(\frac{C}{1-y}\Big)^{1/\alpha} \quad \text{for } y \in (0,1).$$

From this we get, by (13) together with (20) and (21) below,

$$\begin{aligned}
&\sqrt{n}\,\mathbf{E}_{Y_n}\left\{(Q(Y_n)-Q(p))\,I_{Y_n\in(0,\varepsilon)\cup(1-\varepsilon,1)}\right\}\\
&\quad\le \sqrt{n}\int_{y\in(0,\varepsilon)\cup(1-\varepsilon,1)} F^{-1}(y)\frac{y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1}}{B\left(p(n+1),(1-p)(n+1)\right)}dy\\
&\qquad + |Q(p)|\sqrt{n}\,\mathbf{P}_{Y_n}\left\{|Y_n-p|>\varepsilon\right\}\\
&\quad\le \frac{2C^{1/\alpha}\sqrt{n}\,\varepsilon^{(n+1)(p\wedge(1-p))-1/\alpha}}{((n+1)(p\wedge(1-p))-1/\alpha)\,B\left(p(n+1),(1-p)(n+1)\right)}\\
&\qquad + |Q(p)|\sqrt{\mathbf{E}_{Y_n}\left\{\left(\frac{\sqrt{n}\,(Y_n-p)}{\varepsilon f(Q(p))}\right)^2\right\}\mathbf{P}_{Y_n}\{|Y_n-p|>\varepsilon\}}\\
&\quad\overset{d}{\to} 0 \quad \text{as } n\to\infty,
\end{aligned} \tag{14}$$

for $\varepsilon>0$ sufficiently small. Now, putting together (10), (12) and (14), we see that the second term on the right hand side of (9) goes to zero in distribution as $n\to\infty$.

To deal with the first term on the right hand side of (9), we write

$$\begin{aligned}
\sqrt{n}\,\mathbf{E}_{Y_n}\left\{\hat{Q}_n(Y_n)-Q(Y_n)\right\} &= \sqrt{n}\,\mathbf{E}_{Y_n}\left\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\right\}\\
&\quad+\sqrt{n}\,\mathbf{E}_{Y_n}\left\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{Y_n\in[\varepsilon,p-\varepsilon)\cup(p+\varepsilon,1-\varepsilon]}\right\}\\
&\quad+\sqrt{n}\,\mathbf{E}_{Y_n}\left\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{Y_n\in(0,\varepsilon)\cup(1-\varepsilon,1)}\right\}.
\end{aligned} \tag{15}$$

Letting $\{\tilde{X}\}_{i=1}^n$ denote the truncated observations

$$\tilde{X}_i=\begin{cases} Q(p-2\varepsilon) & \text{for} \quad X_i\in(-\infty,Q(p-2\varepsilon))\\ X_i & \text{for} \quad X_i\in[Q(p-2\varepsilon),Q(p+2\varepsilon)]\\ Q(p+2\varepsilon) & \text{for} \quad X_i\in(Q(p+2\varepsilon),\infty)\end{cases} \quad \text{for } \varepsilon>0 \text{ small enough,}$$

and $\tilde{Q}$ and $\tilde{Q}_n$ the corresponding quantile function and empirical quantile process, respectively, the fact that $\hat{Q}_n(p)\to Q(p)$ a.s. as $n\to\infty$ for $p\in(0,1)$ gives

$$\sqrt{n}\,\mathbf{E}_{Y_n}\left\{I_{\tilde{Q}_n(Y_n)\neq\hat{Q}_n(Y_n)}I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\right\}\le\sqrt{n}\left(I_{\hat{Q}_n(p-\varepsilon)<Q(p-2\varepsilon)}+I_{\hat{Q}_n(p+\varepsilon)>Q(p+2\varepsilon)}\right)\to 0 \tag{16}$$

a.s. as $n\to\infty$. From this together with Lemma 3.1, we get that

$$\begin{aligned}
&\left|\sqrt{n}\,\mathbf{E}_{Y_n}\left\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{\tilde{Q}_n(Y_n)\neq\hat{Q}_n(Y_n)}I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\right\}\right|\\
&\le \sup_{u\in(0,1)}\frac{u^{1/\alpha}(1-u)^{1/\alpha}|\hat{Q}_n(u)-Q(u)|}{(p-\varepsilon)^{1/\alpha}(1-p-\varepsilon)^{1/\alpha}}\sqrt{n}\,\mathbf{E}_{Y_n}\left\{I_{\tilde{Q}_n(Y_n)\neq\hat{Q}_n(Y_n)}I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\right\}\\
&\overset{d}{\to} 0 \quad \text{as } n\to\infty.
\end{aligned} \tag{17}$$

And so the mentioned first term is asymptotically equivalent in distribution to

$$\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{\tilde{Q}_n(Y_n)=\hat{Q}_n(Y_n)}I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\Big\}$$
$$=\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\tilde{Q}_n(Y_n)-\tilde{Q}(Y_n))I_{\tilde{Q}_n(Y_n)=\hat{Q}_n(Y_n)}I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\Big\}.$$

Here in turn a straightforward modification of the arguments (16) and (17) give

$$\Big|\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\tilde{Q}_n(Y_n)-\tilde{Q}(Y_n))I_{\tilde{Q}_n(Y_n)\neq\hat{Q}_n(Y_n)}I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\Big\}\Big| \stackrel{d}{\to} 0 \quad \text{as } n\to\infty,$$

so that the first term is asymptotically equivalent in distribution to

$$\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\tilde{Q}_n(Y_n)-\tilde{Q}(Y_n))I_{Y_n\in[p-\varepsilon,p+\varepsilon]}\Big\}.$$

On this, finally, we may use a straightforward modification of the argument(19) below, together with (5), to see that the first term on the right hand side of (15) is asymptotically equivalent in distribution with

$$\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\tilde{Q}_n(Y_n)-\tilde{Q}(Y_n))\Big\} \stackrel{d}{\to} \mathrm{N}\bigg(0,\frac{p(1-p)}{f(Q(p))^2}\bigg) \quad \text{as } n\to\infty \text{ and } \varepsilon\downarrow 0, \tag{18}$$

in that order. As this is the desired central limit, (9) and (10) show that it is enough to prove that the two terms to the right on the right hand side of (15) go to zero as $n\to\infty$, for $\varepsilon>0$ small enough.

For the second term on the right hand side of (15), we have, by arguing as for (17),

$$\begin{aligned}&\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{Y_n\in[\varepsilon,p-\varepsilon)\cup(p+\varepsilon,1-\varepsilon]}\Big\}\\ &\quad\le \sup_{u\in(0,1)}\frac{u^{1/\alpha}(1-u)^{1/\alpha}|\hat{Q}_n(u)-Q(u)|}{\varepsilon^{1/\alpha}(1-\varepsilon)^{1/\alpha}}\;\sqrt{n}\,\mathbf{E}_{Y_n}\big\{I_{Y_n\in[\varepsilon,p-\varepsilon)\cup(p+\varepsilon,1-\varepsilon]}\big\}\\ &\quad\stackrel{d}{\to} 0 \quad \text{as } n\to\infty.\end{aligned} \tag{19}$$

For the third term on the right hand side of (15), notice that by Lemma 3.1,

$$-\Big(\frac{C_n}{y}\Big)^{1/\alpha} < |\hat{Q}_n(y)-Q(y)| < \Big(\frac{C_n}{1-y}\Big)^{1/\alpha} \quad \text{for } y\in(0,1),$$

where $C_n>0$ is a random variable such that $C_n\to 0$ a.s. as $n\to\infty$. And so

$$\begin{aligned}&\Big|\sqrt{n}\,\mathbf{E}_{Y_n}\Big\{(\hat{Q}_n(Y_n)-Q(Y_n))I_{Y_n\in(0,\varepsilon)\cup(1-\varepsilon,1)}\Big\}\Big|\\ &\le \frac{C_n^{1/\alpha}}{B\left(p(n+1),(1-p)(n+1)\right)}\bigg(\int_0^\varepsilon y^{(n+1)p-1-1/\alpha}dy+\int_{1-\varepsilon}^1(1-y)^{(n+1)(1-p)-1-1/\alpha}dy\bigg)\\ &\le \frac{2C_n^{1/\alpha}\varepsilon^{(n+1)(p\wedge(1-p))-1/\alpha}}{((n+1)(p\wedge(1-p))-1/\alpha)\,B\left(p(n+1),(1-p)(n+1)\right)}.\end{aligned} \tag{20}$$

Here, by Stirling's formula [see e.g., Erdélyi, Magnus, Oberhettinger and Tricomi (1953), Equation 1.18.2], together with routine algebra, we have

$$\begin{aligned}&\frac{\varepsilon^{(n+1)(p\wedge(1-p))-1/\alpha}}{((n+1)(p\wedge(1-p))-1/\alpha)\,B\,(p(n+1),(1-p)(n+1))}\\ &\sim\frac{\varepsilon^{(n+1)(p\wedge(1-p))-1/\alpha}}{\sqrt{2\pi p(1-p)(n+1)}\,(p\wedge(1-p))\,p^{(n+1)p}(1-p)^{(n+1)(1-p)}}\to 0\quad\text{as } n\to\infty,\end{aligned}\tag{21}$$

for $\varepsilon > 0$ small enough. Putting this together with (20), we see that also the third term on the right hand side of (15) goes to zero in distribution as $n \to \infty$, as required. □

Proposition 3.3. *Let F be an absolutely continuous distribution function with a strictly positive probability density function f, such that (7) holds with $\alpha \geq 1$. The HD^{HD} estimator satisfies the same central limit theorem as do $\hat{Q}$ and HD:*

$$\sqrt{n}\,\left(HD_n^{HD}(p)-Q(p)\right)\overset{d}{\to}N\!\left(0,\frac{p(1-p)}{f(Q(p))^2}\right)\quad \textit{as } n\to\infty \textit{ for } p\in(0,1).$$

Proof. Keeping the notation from the proof of Proposition 3.2, and letting $Z_n(z)$ denote a beta distributed random variable with parameters $(n+1)z$ and $(n+1)(1-z)$ for $z \in (0,1)$, that is independent of Y_n, we have in the fashion of (9),

$$\begin{aligned}\sqrt{n}\left(\mathrm{HD}_n^{\mathrm{HD}}(p)-Q(p)\right)&=\sqrt{n}\,\mathbf{E}_{Y_n}\left\{\mathrm{HD}_n(Y_n)-Q(p)\right\}\\&=\sqrt{n}\,\mathbf{E}_{Y_n}\left\{T_{n,p}(\hat{Q}_n(Y_n))-Q(p)\right\}\\&=\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\hat{Q}_n(Z_n(Y_n))-Q(p)\right\}\\&=\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\hat{Q}_n(Z_n(Y_n))-Q(Z_n(Y_n))\right\}\\&\quad+\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{Q(Z_n(Y_n))-Q(p)\right\}.\end{aligned}\tag{22}$$

To deal with the second term on the right hand side of (22), we notice that, by the functional central limit theorem (4), applied to the empirical process $\{Z_n(z)-z\}_{z\in(0,1)}$, together with (13) [observing that $|Z_n(p)-p| \leq 1$],

$$\begin{aligned}&\sqrt{n}\;|(Z_n(Y_n)-Y_n)-(Z_n(p)-p)|\\ &\leq \sup_{y\in[p-\varepsilon,p+\varepsilon]}\sqrt{n}\left|\sqrt{n}\,(Z_n(y))-y)-\sqrt{n}\,(Z_n(p)-p)\right|+2\sqrt{n}\,I_{|Y_n-p|>\varepsilon}\overset{d}{\to}0\quad\text{as } n\to\infty.\end{aligned}$$

From this it follows that

$$\sqrt{n}\;(Z_n(Y_n)-p)=\sqrt{n}\;((Z_n(Y_n)-Y_n)-(Z_n(p)-p))+\sqrt{n}\;((Z_n(p)-p)+(Y_n-p))$$

is asymptotically equal in distribution to $\sqrt{n}\ ((Z_n(p) - p) + (Y_n - p))$, so that, by (8),

$$\sqrt{n}\ (Z_n(Y_n)) - p) \xrightarrow{d} \mathrm{N}\,(0, 2p(1-p)) \quad \text{as } n \to \infty. \tag{23}$$

Hence an inspection of the proof of Proposition 3.2 gives [cf. (10) and (12)]

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{(Q(Z_n(Y_n)) - Q(p))\, I_{|Z_n(Y_n)-p|\leq\varepsilon}\Big\} \xrightarrow{d} 0 \quad \text{as } n \to \infty \text{ and } \varepsilon \downarrow 0,$$

in that order, and

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{(Q(Z_n(Y_n)) - Q(p))\, I_{Z_n(Y_n)\in[\varepsilon,p-\varepsilon)\cup(p+\varepsilon,1-\varepsilon]}\Big\} \xrightarrow{d} 0 \quad \text{as } n \to \infty.$$

To show that the second term on the right hand side of (22) is asymptotically neglible as $n \to \infty$ (and $\varepsilon \downarrow 0$), it therefore only remains to show that

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{(Q(Z_n(Y_n)) - Q(p))\, I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}\Big\} \xrightarrow{d} 0 \quad \text{as } n \to \infty.$$

As an inspection of (14) shows that

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{Q(p) I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}\Big\} \xrightarrow{d} 0 \quad \text{as } n \to \infty,$$

this in turn will follow if we can prove that

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{Q(Z_n(Y_n)) I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}\Big\} \xrightarrow{d} 0 \quad \text{as } n \to \infty. \tag{24}$$

To prove (24) we observe that, by Stirling's formula,

$$\begin{aligned}
&\frac{1}{B((n+1)p, (n+1)(1-p))B((n+1)y, (n+1)(1-y))} \\
&\leq C(n+1)\exp\{-(n+1)\,(y\ln(y) + p\ln(p) + (1-y)\ln(1-y) + (1-p)\ln(1-p))\} \\
&\leq C(n+1)\exp\{C(n+1)\}
\end{aligned} \tag{25}$$

for $n > 0$ and $y, p \in (0,1)$, for some constant $C > 0$. It follows that

$$\begin{aligned}
&\Big|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{Q(Z_n(Y_n)) I_{Z_n(Y_n)\in(0,\varepsilon)} I_{Y_n\in(0,\epsilon)\cup(1-\epsilon,1)}\Big\}\Big| \\
&= \Big|\sqrt{n}\Big(\int_{y=0}^{y=\epsilon} + \int_{y=1-\epsilon}^{1}\Big)\int_{z=-\infty}^{z=Q(\varepsilon)} \\
&\qquad \frac{zf(z)F(z)^{(n+1)y-1}(1-F(z))^{(n+1)(1-y)-1}y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1}}{B((n+1)y, (n+1)(1-y))B((n+1)p, (n+1)(1-p))}dzdy\Big| \\
&\leq C\sqrt{n}\,(n+1)\exp\{C(n+1)\}\Big(\int_{\mathbb{R}}|z|f(z)dz\Big)\Big(\int_0^{\epsilon}\frac{\varepsilon^{(n+1)y}y^{(n+1)p-1}}{\varepsilon}dy
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad + \int_{1-\epsilon}^{1} \frac{\varepsilon^{(n+1)y}(1-y)^{(n+1)p-1}}{\varepsilon} dy \Big) \\
&\le C\sqrt{n}\Big(\int_{\mathbb{R}} |z| dF(z)\Big)\Big(\frac{\exp\{(\ln(\epsilon)p + C)(n+1)\}}{p\epsilon} + \frac{\exp\{(\ln(\epsilon)(1-p) + C)(n+1)\}}{(1-p)\epsilon}\Big) \\
&\to 0 \quad \text{as } n\to\infty \text{ for } \epsilon > 0 \text{ small enough.}
\end{aligned} \tag{26}
$$

By a symmetric argument, we get

$$
\Big|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{Q(Z_n(Y_n))I_{Z_n(Y_n)\in(1-\varepsilon,1)}I_{Y_n\in(0,\epsilon)\cup(1-\epsilon,1)}\Big\}\Big| \to 0 \quad \text{as } n\to\infty. \tag{27}
$$

Moreover, we readily get in a similar fashion, making use of (25) again,

$$
\begin{aligned}
&\Big|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{Q(Z_n(Y_n))I_{Z_n(Y_n)\in(0,\varepsilon)}I_{Y_n\in(\epsilon,1-\epsilon)}\Big\}\Big| \\
&\qquad \le \sqrt{n}\,C(n+1)\exp\{C(n+1)\}\int_{y=\epsilon}^{y=1-\epsilon}\int_{z=-\infty}^{z=Q(\varepsilon)} \frac{|z|f(z)F(z)^{(n+1)\epsilon-1}}{y(1-y)} dz \\
&\qquad \le \frac{\sqrt{n}\,C(n+1)}{\epsilon^2}\Big(\int_{\mathbb{R}}|z|f(z)dz\Big)\exp\{C(n+1)\}\,\varepsilon^{(n+1)\epsilon} \to 0 \quad \text{as } n\to\infty
\end{aligned} \tag{28}
$$

for $\varepsilon > 0$ small enough, and by the symmetric argument

$$
\Big|\mathbf{E}_{Y_n,Z_n}\Big\{Q(Z_n(Y_n))I_{Z_n(Y_n)\in(1-\varepsilon,1)}I_{Y_n\in(\epsilon,1-\epsilon)}\Big\}\Big| \to 0 \quad \text{as } n\to\infty. \tag{29}
$$

Putting (26)-(29) together, we get the desired asymptotics (24).

For the first term on the right hand side of (22), notice that, because of (23), by inspection of the proof of Proposition 3.2, the following versions of (18) and (19) hold:

$$
\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{\Big(\hat{Q}_n(Z_n(Y_n)) - Q(Z_n(Y_n))\Big)I_{Z_n(Y_n)\in[p-\varepsilon,p+\varepsilon]}\Big\} \xrightarrow{d} \mathrm{N}\Big(0, \frac{p(1-p)}{f(Q(p))^2}\Big)
$$

as $n\to\infty$ and $\varepsilon\downarrow 0$, in that order, and

$$
\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{\Big(\hat{Q}_n(Z_n(Y_n)) - Q(Z_n(Y_n))\Big)I_{Z_n(Y_n)\in[\varepsilon,p-\varepsilon)\cup(p+\varepsilon,1-\varepsilon]}\Big\} \xrightarrow{d} 0 \quad \text{as } n\to\infty.
$$

As the second term on the right hand side of (22) has been shown to be asymptotically neglible as $n\to\infty$ (and $\varepsilon\downarrow 0$), it therefore only remains to show that

$$
\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{\Big(\hat{Q}_n(Z_n(Y_n)) - Q(Z_n(Y_n))\Big)I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}\Big\} \xrightarrow{d} 0 \quad \text{as } n\to\infty
$$

for $\varepsilon > 0$ small enough. In view of (24), this in turn will follow if we prove that

$$
\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{\hat{Q}_n(Z_n(Y_n))I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}\Big\} \xrightarrow{d} 0 \quad \text{as } n\to\infty. \tag{30}
$$

To prove (30), we use the approach in (26)-(29) again. When estimating

$$\left|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\hat{Q}_n(Z_n(Y_n))I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}I_{Y_n\in(0,\epsilon)\cup(1-\epsilon,1)}\right\}\right|$$

and

$$\left|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\hat{Q}_n(Z_n(Y_n))I_{Z_n(Y_n)\in(0,\varepsilon)\cup(1-\varepsilon,1)}I_{Y_n\in(\epsilon,1-\epsilon)}\right\}\right|$$

in that fashion, the only changes in (26) and (28), are that the probability distribution function $F(z)$ is replaced with the empirical distribution function $F_n(z)$, the probability density function $f(z)$ with the empirical Stieltjes measure $dF_n(z)$, and the quantile function $Q(p)$ with the empirical quantile process $\hat{Q}_n(p)$. And so it is readily seen that the arguments from (26)-(29) carry over, with these changes only, to yield the desired asymptotics (30), using the strong law of large numbers to conclude from (7) that

$$\int_{\mathbb{R}} |x|dF_n(x) \to \int_{\mathbb{R}} |x|dF(x) \text{ a.s. as } n\to\infty. \qquad \Box \tag{31}$$

To deal with the $\mathrm{HD}^{\mathrm{Kernel}}$ estimator, recall that a function $g:\mathbb{R}\to[0,\infty)$ is called *almost increasing* if $g(x)\geq cg(y)$ for $x\geq y$, for some constant $c>0$, and *almost decreasing* if $g(x)\leq Cg(y)$ for $x\geq y$, for some constant $C<\infty$, see e.g., Bingham, Goldie and Teugels (1987), p. 72. Motivated by this terminology, we call g *almost unimodal* if there exists a mode $m\in\mathbb{R}$ such that

$$g(x)\leq Cg(y) \quad \text{for } x\leq y\leq m \quad \text{and} \quad g(x)\leq Cg(y) \quad \text{for } m\leq y\leq x.$$

Proposition 3.4. *Let F be an absolutely continuous distribution function with a strictly positive and probability density function f, that is differentiable with a bounded derivative. Provided that (7) holds with $\alpha\geq 1$, that*

$$\lim_{n\to\infty} n^{1/4}h(n)=0 \quad \textit{and} \quad \lim_{n\to\infty} n^{\rho}h(n)=\infty \quad \textit{for some } \rho>0, \tag{32}$$

and that K is almost unimodal with

$$\lim_{n\to\infty}\sqrt{n}\int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} x^2K_{h(n)}(x)dx=0 \quad \textit{for } \varepsilon>0, \tag{33}$$

the HD^{Kernel} estimator satisfies the same central limit theorem as do $\hat{Q}$, HD, and HD^{HD}:

$$\sqrt{n}\,\left(HD_n^{Kernel}(p)-Q(p)\right)\xrightarrow{d} N\!\left(0,\frac{p(1-p)}{f(Q(p))^2}\right) \quad \textit{as } n\to\infty \textit{ for } p\in(0,1).$$

Proof. Defining $\hat{Q}_n(p) = 0$ for $p \in (0,1)$, and letting Z_n denote a random variable, independent of Y_n, with probability density function $K_h = K_{h(n)}$, we may express, in the fashion of (22),

$$\begin{aligned} &\sqrt{n}\left(\mathrm{HD}_n^{\mathrm{Kernel}}(p) - Q(p)\right) \\ &= \sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\hat{Q}_n(Y_n+Z_n) - Q(Y_n+Z_n)\right\} + \sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{Q(Y_n+Z_n) - Q(p)\right\}. \end{aligned} \tag{34}$$

Here, for the the second term on the right hand side, by (32), we have to work more than for (10), because Z_n is $o(n^{-1/4})$ rather than $O(n^{-1/2})$ as $Y_n - p$: By Taylor expansion, together with the fact that $\mathbf{E}\{Y_n - p\} = \mathbf{E}\{Z_n\} = 0$, (8) and (33) readily give

$$\begin{aligned} &\left|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{(Q(Y_n+Z_n) - Q(p))\, I_{|Y_n-p|\le\varepsilon} I_{|Z_n|\le\varepsilon}\right\}\right| \\ &= \left|\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\left(Q'(p)(Y_n+Z_n-p) + \int_{q=p}^{q=Y_n+Z_n}\int_{r=p}^{r=q} Q''(r)\,dr\,dq\right) I_{|Y_n+Z_n-p|\le 2\varepsilon}\right\}\right| \\ &\le \sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\frac{|Y_n+Z_n-p|}{f(Q(p))} I_{|Y_n+Z_n-p|>2\varepsilon} + (Y_n+Z_n-p)^2 \sup_{q\in[p-2\varepsilon,p+2\varepsilon]}\left|\frac{f'(Q(q))}{f(Q(q))^3}\right|\right\} \\ &\le \sqrt{\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\frac{(Y_n+Z_n-p)^2}{f(Q(p))}\right\}\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\frac{(Y_n+Z_n-p)^2}{4\varepsilon^2}\right\}} \\ &\qquad + \sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{(Y_n+Z_n-p)^2\right\} \sup_{q\in[p-2\varepsilon,p+2\varepsilon]}\left|\frac{f'(Q(q))}{f(Q(q))^3}\right| \\ &\xrightarrow{d} 0 \quad \text{as } n\to\infty. \end{aligned} \tag{35}$$

Hence, as (13) and (33) show that

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{|Q(p)| I_{|Y_n-p|>\varepsilon}\right\} \xrightarrow{d} 0 \quad \text{and} \quad \sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{|Q(p)| I_{|Z_n|>\varepsilon}\right\} \xrightarrow{d} 0 \quad \text{as } n\to\infty,$$

in order to show that

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{Q(Y_n+Z_n) - Q(p)\right\} \xrightarrow{d} 0 \quad \text{as } n\to\infty, \tag{36}$$

it remains to prove

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{|Q(Y_n+Z_n)| I_{|Y_n-p|>\varepsilon}\right\} \xrightarrow{d} 0 \quad \text{as } n\to\infty, \tag{37}$$

and

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{|Q(Y_n+Z_n)| I_{|Z_n|>\varepsilon}\right\} \xrightarrow{d} 0 \quad \text{as } n\to\infty. \tag{38}$$

To prove (37), notice that, since Y_n has a distribution that is unimodal with mode $((n+1)p-1)/(n-1)$, and since

$$\lim_{n\to\infty} n^{-\rho} \sup_{x\in\mathbb{R}} K_{h(n)}(x) < \infty$$

by (32), we have

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{|Q(Y_n+Z_n)|I_{|Y_n-p|>\varepsilon}\Big\}$$
$$\le \sqrt{n}\bigg(\int_{y=0}^{y=p-\varepsilon}+\int_{y=p+\varepsilon}^{1}\bigg)\int_{z=-\infty}^{z=\infty}\frac{|z|f(z)K_{h(n)}(F(z)-y)y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1}}{B((n+1)p,(n+1)(1-p))}\,dzdy$$
$$\le \sqrt{n}\bigg(\int_{\mathbb{R}}|z|f(z)dz\bigg)\sup_{x\in\mathbb{R}}K_{h(n)}(x)$$
$$\times\frac{(p-\varepsilon)^{(n+1)p-1}(1-(p-\varepsilon))^{(n+1)(1-p)-1}+(p+\varepsilon)^{(n+1)p-1}(1-(p+\varepsilon))^{(n+1)(1-p)-1}}{B((n+1)p,(n+1)(1-p))}$$
$$\to 0\quad\text{as } n\to\infty: \tag{39}$$

Here we used Stirling's formula in the fashion of (25), to see that

$$n^{\rho}\frac{(p-\varepsilon)^{(n+1)p-1}(1-(p-\varepsilon))^{(n+1)(1-p)-1}+(p+\varepsilon)^{(n+1)p-1}(1-(p+\varepsilon))^{(n+1)(1-p)-1}}{B((n+1)p,(n+1)(1-p))}$$
$$\le C\sqrt{n+1}\,\frac{\exp\{-(n+1)\left(p\ln(p/(p-\varepsilon))+(1-p)\ln((1-p)/(1-p+\varepsilon))\right)\}}{(p-\varepsilon)(1-p+\varepsilon)}$$
$$+C\sqrt{n+1}\,\frac{\exp\{-(n+1)\left(p\ln(p/(p+\varepsilon))+(1-p)\ln((1-p)/(1-p-\varepsilon))\right)\}}{(p+\varepsilon)(1-p-\varepsilon)}$$
$$\to 0\quad\text{as } n\to\infty,$$

because the function

$$g(\varepsilon)=p\ln(p/(p-\varepsilon))+(1-p)\ln((1-p)/(1-p+\varepsilon))$$

satisfies $g(0)=g'(0)=0$ and $g''(\varepsilon)>0$.

For the proof of (38), it is enough to notice that, in a way similar to (39), by (33) together with the almost unimodality of K,

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\Big\{|Q(Y_n+Z_n)|I_{|Z_n|>\varepsilon}\Big\}\le\sqrt{n}\bigg(\int_{\mathbb{R}}|z|f(z)dz\bigg)\sup_{|x|>\varepsilon}K_{h(n)}(x)\to 0\quad\text{as } n\to\infty.$$

For the first term on the right hand side of (34), by inspection of the proof of Proposition 3.2, and in the fashion of the proof of Proposition 3.3, we have [cf. (18)]

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{\Big(\hat{Q}_n(Y_n+Z_n)-Q(Y_n+Z_n))\Big)I_{Y_n\in[p-\varepsilon,p+\varepsilon]}I_{|Z_n|\le\varepsilon}\right\}\xrightarrow{d}\mathrm{N}\bigg(0,\frac{p(1-p)}{f(Q(p))^2}\bigg)$$

as $n\to\infty$ and $\varepsilon\downarrow 0$, in that order. And so, by (37) and (38), it remains to show that

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{|\hat{Q}_n(Y_n+Z_n)|I_{|Y_n-p|>\varepsilon}\right\}\xrightarrow{d}0\quad\text{as } n\to\infty, \tag{40}$$

and

$$\sqrt{n}\,\mathbf{E}_{Y_n,Z_n}\left\{|\hat{Q}_n(Y_n+Z_n)|I_{|Z_n|>\varepsilon}\right\} \xrightarrow{d} 0 \quad \text{as } n\to\infty. \tag{41}$$

However, this is done in the same way as the proof of (37) and (38). The only changes are that the probability distribution function $F(z)$ is replaced with the empirical distribution function $F_n(z)$, the density function $f(z)$ with the empirical Stieltjes measure $dF_n(z)$, and the quantile function $Q(p)$ with the empirical quantile process $\hat{Q}_n(p)$. With these changes, the arguments for (37) and (38) carry over in an obvious fashion to show (40) and (41), using the strong law of large numbers to conclude (31) from (7). □

4 Bias Correction

As our L-estimator uses more orders statistics than the sample quantile, it is natural to expect that it gives a larger bias. However, one could try to lower the bias by subtracting a bootstrap estimate of it. Normally, one would have to resample to do this. But in our case one can calculate an exact bootstrap mean, as stated above, using that the Harrell-Davis estimator is the exact bootstrap mean. Hence we would have the following bias correction for an L-estimator L_n:

$$\mathbf{E}_{\text{BOOT}}\{L_n(p)\} - \hat{Q}_n(p),$$

where

$$\mathbf{E}_{\text{BOOT}}\{L_n(p)\} = \sum_{i=1}^{n} w_i\,\mathbf{E}_{F_n}\{X_{(i)}\} = \sum_{i=1}^{n} w_i\,\mathrm{HD}_n(i/(n+1)).$$

However, for a large sample size n, this could be quite computationally expensive.

5 Selecting the Bandwidth

According to Sheather and Marron (1990), the asymptotically optimal bandwidth h, in mean square error sense, MSE, for a kernel estimator, when F is symmetric, is given by

$$h_{\text{opt}}(n) = \alpha(K)\beta(Q)n^{-1/3},$$

where

$$\beta(Q) = \left(\frac{Q'(p)}{Q''(p)}\right)^{2/3} \quad \text{and} \quad \alpha(K)^3 = \frac{2\int_{-\infty}^{\infty} uK(u)(\int K)(u)du}{\left(\int_{-\infty}^{\infty} u^2K(u)du\right)^2},$$

with $\int K$ denoting a primitive function of K.

It has been argued that one should assume normal distribution to calculate $\beta(Q)$, when one do not possess a priori information on the probability distribution of the sample, see Sheather and Marron (1990), p. 415. However, by means of common statistical procedures, one can usually find a probability distribution that fits the data at hand better than do the normal distribution. We will work in this way, trying to find a better a priori guess of the distribution at hand, than simply assuming normal distribution, in order to get a better choice of the bandwidth.

Notice that with the described method, the estimated optimal bandwidth can force the kernel estimator to give positive weight to empirical quantiles $\hat{Q}_n(p)$ from outside the interval $p \in (0, 1)$. This problem is taken care of in the following crude way: If an estimator $\hat{h}$ of the optimal bandwidth has the property that $\hat{h}+p > 0.99$ or $p-\hat{h} < 0.01$, then we set $\hat{h} = (1-p)/2$ and $\hat{h} = p/2$, respectively.

6 Comparison with $\hat{Q}_n(p)$

As have been shown in previous sections, our new estimators behaves asymptotically as the empirical quantile, which motivates using them for a large sample, or perhaps not to use them, as the empirical quantile is a simpler estimator. To investigate how our estimators behave for a finite sample, we conduct simulations studies. Similar methods on comparing estimators can be found in Sheather and Marron (1990) and Yang (1985).

We will compare our proposed modified HD estimators with the estimator $\hat{Q}_n(p)$, in the sense of relative mean square error, ΔMSE, as it is a natural trade off between bias and variance: If $S_n(p)$ is an estimator of $Q(p)$, that error is given by

$$\Delta\mathrm{MSE}(S_n(p)) = \frac{\mathrm{MSE}(\hat{Q}_n(p))}{\mathrm{MSE}(S_n(p))} = \frac{\mathbf{E}\{(\hat{Q}_n(p) - Q(p))^2\}}{\mathbf{E}\{(S_n(p) - Q(p))^2\}}.$$

Of course, if the relative mean square error is larger then 1, then the new estimator is better then the empirical quantile.

As the bandwidth has to be estimated, the relative mean square error cannot be calculated exactly. Rather, it has to be estimated. This is done by Monte Carlo computer simulation, where we generate samples of random variables, and then estimate ΔMSE.

We will try four different distributions: A standard normal distribution, a Student t distribution with 4 degrees of freedom, a standard lognormal distribution, and a standard exponential distribution, with sample sizes $n = 25$, 50, 100 and 1000, for the 0.05, 0.1, 0.25, 0.4, 0.6, 0.75, 0.9 and 0.95 quantiles for standard lognormal and exponential distribution. For reasons of symmetry, we only consider the 0.05, 0.1, 0.2, 0.3, 0.4 and 0.45 quantiles for the standard normal and the Student t distribution.

In the kernel based methods, we will use the normal distribution kernel

$$K(u) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-u^2/2} \quad \text{with} \quad \alpha(K) = \frac{1}{2\sqrt{\pi}}.$$

For estimation of $\beta(Q)$, we generously assume that we know the distribution family, and only have to estimate its parameters, making it straightforward to calculate $\beta(Q)$.

7 Numerical Results

Due to computational issues, it is hard to get confidence interval for the relative MSE for the $\mathrm{HD}^{\mathrm{Kernel}}$ and the K estimators. This is so because for each sample one has to calculate the kernel. However, initial tests on the HD and $\mathrm{HD}^{\mathrm{HD}}$ estimators, where confidence interval were calculated, showed that one needs at about 10000 Monte Carlo simulations to get sufficient small confidence interval, ± 0.02. Compare this to Sheather and Marron (1990), who used 1000 simulations: We did 10000 Monte Carlo simulations.

During the initial test of our estimators, we discovered that the bias correction indeed lowered the bias, but the introduction of $\hat{Q}_n(p)$ increased the variance so much that the overall performance were less good in MSE sense. Therefore we did not do bias corrections in the simulations studies.

The complete numerical results of our simulation study is presented in Appendix A.

Examing the results of the simulation study we see that the $\mathrm{HD}^{\mathrm{HD}}$ estimator performed between 10% to 70% better than the empirical quantile for the exponential and lognormal distributions, for probabilities under 0.6, and for all sample sizes. It performed up to 20% better in the center of the normal and Student t distributions. Further, one can see that the $\mathrm{HD}^{\mathrm{HD}}$ estimator had difficulties with a heavy tail.

The HD estimator also performed well, and similarly to the $\mathrm{HD}^{\mathrm{HD}}$ estimator. Looking in detail, the HD estimator performed better in the tails, but less good in the center of a distribution, than did the $\mathrm{HD}^{\mathrm{HD}}$ estimator.

Also the $\mathrm{HD}^{\mathrm{Kernel}}$ and K estimators performed similarly. They behaved really well for the standard normal distribution, for all quantiles. For the Student t distribution they behaved less good than the $\mathrm{HD}^{\mathrm{HD}}$ and HD estimatots in the center, but better in the tail of the distribution. However, in the tails the performance was less good than the empirical quantile. For the non symmetric distributions, the $\mathrm{HD}^{\mathrm{Kernel}}$ and K estimators performed less good than the $\mathrm{HD}^{\mathrm{HD}}$ and HD estimators, except in the tails.

We can conclude that if one wants to estimate a quantile that is not in the tail, then the $\mathrm{HD}^{\mathrm{HD}}$ estimator is a good choice, that is slightly better than the HD estimator. This is so because not only because it performs well, but also because one does not have to select a bandwidth, which one has to do with the K estimator. For the tails the conclusion is that one should use the empirical quantile estimator. Also, when dealing with a normal distributed sample, one could use the $\mathrm{HD}^{\mathrm{Kernel}}$ and K estimators.

References

Bingham, N.H., Goldie, C.M. and Teugels J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

David, H.A. (1981). *Order Statistics (2nd Ed.).* John Wiley, New York.

Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F.G. (1953). *Higher Transcendental Functions, Vol. I.* McGraw-Hill, New York.

Harrell, F.E. and Davis, C.E. (1982). A new disribution-free quantile estimator. *Biometrika* **69** 635-640.

Hutson, A.D. and Ernst, M.D. (2000). The exact bootstrap mean and variance of an L-estimator. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 89-94.

Lehmann, R.J. (1983). *Theory of Point Estimation.* John Wiley, New York.

Mason, D.M. (1982). Some characterizations of almost sure bounds for weighted multidimensional empirical distributions and a Glivenko-Cantelli theorem for sample quantiles. *Z. Wahrsch. Verw. Gebiete* **59** 505-513.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Sheather, S.J. and Marron, J.S. (1990). Kernal quantile estimators. *J. Amer. Statist.*

Assoc. **85** 410-416.

van der Vaart, A.W. and Weller, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag, New York.

Yang, S.S. (1985). A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc.* **80** 1004-1011.

Yoshizawa, C.N., Sen, P.K. and Davis C.E. (1985). Asymptotic equivalence of the Harrell-Davis median estimator and the sample median. *Comm. Statist. Theory and Methods* **14**, 2129-2136.

Zelterman, D. (1990). Nonparametric quantile estimation. *J. Statist. Plann. and Inference* **26** 339-352.

Corresponding Author

Erik Brodin

Department of Mathematical Sciences

Chalmers University of Technology

SE-412 96 Göteborg

Sweden

ebrodin@math.chalmers.se

http://www.math.chalmers.se/~brodin

A Appendix: Numerical Results of Simulations

Here we present the corresponding relative mean square error from the different estimators, different distributions, sample sizes and quantiles.

A.1 Normal Distribution

Normal, n=25	0.05	0.1	0.2	0.3	0.4	0.45
HD	1.09	1.19	1.24	1.20	1.24	1.22
HD^{HD}	1.06	1.09	1.24	1.25	1.31	1.30
K	1.22	1.27	1.25	1.24	1.31	1.48
HD^{Kernel}	1.31	1.32	1.25	1.24	1.31	1.48

Normal, n=50	0.05	0.1	0.2	0.3	0.4	0.45
HD	1.12	1.18	1.20	1.18	1.17	1.16
HD^{HD}	0.99	1.12	1.24	1.23	1.23	1.22
K	1.24	1.19	1.20	1.09	1.14	1.44
HD^{Kernel}	1.30	1.19	1.18	1.07	1.13	1.44

Normal, n=100	0.05	0.1	0.2	0.3	0.4	0.45
HD	1.16	1.15	1.13	1.11	1.12	1.11
HD^{HD}	1.06	1.14	1.17	1.15	1.16	1.15
K	1.19	1.19	1.21	1.14	1.04	1.29
HD^{Kernel}	1.19	1.16	1.17	1.11	1.03	1.29

Normal, n=1000	0.05	0.1	0.2	0.3	0.4	0.45
HD	1.07	1.06	1.04	1.04	1.03	1.03
HD^{HD}	1.08	1.07	1.06	1.05	1.05	1.05
K	1.10	1.11	1.12	1.14	1.24	0.98
HD^{Kernel}	1.08	1.10	1.12	1.14	1.24	0.98

A.2 Student t Distribution

Student t, n=25	0.05	0.1	0.2	0.3	0.4	0.45
HD	0.52	0.77	1.02	1.15	1.17	1.18
HD^{HD}	0.47	0.50	0.81	1.11	1.20	1.22
K	0.77	0.93	0.94	1.07	1.12	1.15
HD^{Kernel}	0.86	0.91	0.92	1.11	1.11	1.14

Student t, n=50	0.05	0.1	0.2	0.3	0.4	0.45
HD	0.67	0.91	1.08	1.13	1.16	1.14
HD^{HD}	0.41	0.63	1.01	1.13	1.20	1.18
K	0.89	0.87	0.98	1.02	1.18	1.05
HD^{Kernel}	0.81	0.76	0.91	0.99	1.18	1.05

Student t, n=100	0.05	0.1	0.2	0.3	0.4	0.45
HD	0.85	1.02	1.09	1.11	1.10	1.11
HD^{HD}	0.56	0.90	1.08	1.13	1.13	1.14
K	0.84	0.96	1.03	1.05	1.06	1.14
HD^{Kernel}	0.71	0.82	0.95	1.00	1.06	1.13

Student t, n=1000	0.05	0.1	0.2	0.3	0.4	0.45
HD	1.04	1.04	1.04	1.04	1.04	1.03
HD^{HD}	1.03	1.04	1.06	1.05	1.05	1.04
K	1.02	1.03	1.08	1.07	1.07	1.02
HD^{Kernel}	0.98	1.00	1.07	1.07	1.07	1.02

A.3 Lognormal Distribution

Lognormal, n=25	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.61	1.33	1.29	1.33	1.24	0.83	0.58	0.38
HD^{HD}	1.57	1.44	1.33	1.26	1.04	0.57	0.33	0.33
K	1.45	1.22	0.93	0.91	1.07	0.77	0.79	0.60
HD^{Kernel}	1.30	1.14	0.76	0.74	0.83	0.57	0.66	0.65

Lognormal, n=50	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.37	1.50	1.17	1.21	1.18	1.01	1.12	0.60
HD^{HD}	1.51	1.67	1.21	1.21	1.10	0.87	0.67	0.35
K	1.25	1.50	0.53	1.03	1.07	0.94	1.12	0.82
HD^{Kernel}	1.24	1.46	0.46	0.92	0.92	0.74	0.84	0.71

Lognormal, n=100	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.45	1.30	1.17	1.15	1.13	1.13	1.12	1.11
HD^{HD}	1.61	1.40	1.22	1.17	1.11	1.07	0.89	0.65
K	1.47	1.41	0.43	1.09	1.07	1.08	1.07	1.12
HD^{Kernel}	1.48	1.39	0.39	1.04	1.00	0.96	0.81	0.85

Lognormal, n=1000	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.10	1.07	1.05	1.04	1.04	1.04	1.05	1.07
HD^{HD}	1.14	1.10	1.06	1.05	1.05	1.04	1.05	1.05
K	1.35	1.25	1.09	1.07	1.06	1.04	1.04	1.05
HD^{Kernel}	1.36	1.21	1.09	1.07	1.05	1.03	1.01	1.00

A.4 Exponential Distribution

Exponential, n=25	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.65	1.34	1.33	1.36	1.29	1.03	0.94	0.82
HD^{HD}	1.53	1.43	1.39	1.36	1.23	0.90	0.72	0.78
K	1.49	1.19	0.44	1.01	1.15	0.97	1.12	1.08
HD^{Kernel}	1.22	1.06	0.37	0.85	1.07	0.94	1.19	1.19

Exponential, n=50	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.32	1.54	1.16	1.23	1.19	1.09	1.25	0.92
HD^{HD}	1.42	1.67	1.22	1.26	1.19	1.05	1.01	0.70
K	1.18	1.47	0.55	1.10	1.12	1.02	1.25	1.11
HD^{Kernel}	1.05	1.35	0.47	1.00	1.03	0.94	1.23	1.17

Exponential, n=100	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.55	1.33	1.20	1.15	1.14	1.15	1.19	1.23
HD^{HD}	1.71	1.43	1.25	1.18	1.16	1.15	1.10	0.98
K	1.45	1.35	0.82	1.14	1.12	1.11	1.13	1.24
HD^{Kernel}	1.39	1.27	0.74	1.10	1.08	1.04	1.04	1.22

Exponential, n=1000	0.05	0.1	0.25	0.4	0.6	0.75	0.9	0.95
HD	1.11	1.08	1.04	1.04	1.04	1.05	1.05	1.08
HD^{HD}	1.15	1.11	1.06	1.06	1.05	1.06	1.06	1.08
K	1.19	0.97	1.13	1.08	1.07	1.06	1.06	1.07
HD^{Kernel}	1.16	0.98	1.12	1.08	1.06	1.05	1.04	1.04

PAPER II

On Quantile Estimation by Bootstrap

Erik Brodin

Chalmers University of Technology

January 25, 2005

Abstract

We use exact bootstrap to optimize the weights of an L-estimator for quantiles with respect to the estimated MSE (mean square error). We test the performance of the new estimator by comparing with that of the sample quantile. The new estimator performs better than the sample quantiles in almost every case. However, the gain is only about 5%, in terms of decreased MSE.

Keywords: Bootstrap; L-estimator; Order statistics; Quantile estimation.

1 Introduction

Let $X_1, \ldots, X_n$ be independent and identically distribtuted random variables with probability distribution function F. Let $X_{(1)} \leq \ldots \leq X_{(n)}$ be the corresponding ordered sample. We define the quantile function as the left inverse of F

$$Q(p) = \inf\{x : F(x) \geq p\} \quad \text{for } 0 < p < 1.$$

A basic estimator of $Q(p)$, the p:th quantile, is the p:th sample quantile $\hat{Q}_n(p) = X_{(\lfloor np \rfloor + 1)}$, where $\lfloor x \rfloor$ is the integer part of $x \in \mathbb{R}$. However, it might be possible to improve this estimator by using an L-estimator that averages over the order statistics:

$$L_n(p) = \sum_{i=1}^{n} w_i X_{(i)} \quad \text{where} \quad \sum_{i=1}^{n} w_i = 1. \tag{1}$$

The choice of the weights w_i in (1) is a trade off between bias and variance: The more order statistics that are involved in the L-estimator, the smaller will be the variance of the estimator, but the larger will be the bias, and vice versa for less order statistcs.

In this article, we make a simulation study of how it works to select the weights w_i for the L-estimator (1) by minimizing a bootstrap estimate of the MSE of the estimator. As this gives quite poor control of the bias, we have selected to use a fixed number of weights, as otherwise we get an estimator with small variance, but too large bias. And so we have selected to use 3 non-zero weights $w_{\lfloor np \rfloor}$, $w_{\lfloor np \rfloor+1}$ and $w_{\lfloor np \rfloor+2}$, that should sum up to one and be symmetric $w_{\lfloor np \rfloor} = w_{\lfloor np \rfloor+2}$, to get a small bias.

2 Bootstrap Estimates of MSE

By an optimal L-estimator L_n, we mean an estimator that has the minimal MSE. Hence, we look for the weights w_i in (1) that minimize

$$\begin{aligned}
&\mathrm{MSE}(L_n(p)) \\
&= \mathbf{E}\{(L_n(p) - Q(p))^2\} \\
&= \sum_{i=1}^{n} w_i^2 \mathbf{Var}\{X_{(i)}\} + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} w_i w_j \mathbf{Cov}\{X_{(i)}, X_{(j)}\} + \left(\sum_{i=1}^{n} w_i \mathbf{E}\{X_{(i)}\} - Q(p)\right)^2 .
\end{aligned} \tag{2}$$

To be able to look for optimal weights w_i in (1), we use the bootstrap estimate $\mathrm{MSE}_{\mathrm{Boot}}(L_n(p))$ of the MSE (2), where the unknown quantities $\mathbf{Var}\{X_{(i)}\}$, $\mathbf{Cov}\{X_{(i)}, X_{(j)}\}$, $\mathbf{E}\{X_{(i)}\}$ and $Q(p)$ in (2) have been replaced with their bootstrap estimates $\mathbf{Var}_{\mathrm{Boot}}\{X_{(i)}\}$, $\mathbf{Cov}_{\mathrm{Boot}}\{X_{(i)}, X_{(j)}\}$, $\mathbf{E}_{\mathrm{Boot}}\{X_{(i)}\}$ and $\hat{Q}_n(p)$, respectively.

We calculate the bootstrap estimators required by exact bootstrap, following [1]: According to them, we have

$$\begin{aligned}
\mathbf{E}_{\mathrm{Boot}}\{X_{(r)}\} &= \sum_{i=1}^{n} a_{i,r} X_{(i)}, \\
\mathbf{Var}_{\mathrm{Boot}}\{X_{(r)}\} &= \sum_{i=1}^{n} a_{i,r} (X_{(i)} - \mathbf{E}_{\mathrm{Boot}}\{X_{(r)}\})^2, \\
\mathbf{Cov}_{\mathrm{Boot}}\{X_{(r)}, X_{(s)}\} &= \sum_{j=2}^{n}\sum_{i=1}^{j-1} b_{ij,rs} (X_{(i)} - \mathbf{E}_{\mathrm{Boot}}\{X_{(r)}\})(X_{(j)} - \mathbf{E}_{\mathrm{Boot}}\{X_{(s)}\}) \\
&\quad + \sum_{i=1}^{n} c_{i,rs} (X_{(i)} - \mathbf{E}_{\mathrm{Boot}}\{X_{(r)}\})(X_{(i)} - \mathbf{E}_{\mathrm{Boot}}\{X_{(s)}\}).
\end{aligned}$$

Here, writing β for the incomplete β-function $\beta(p,q;z) = \int_0^z x^{p-1}(1-x)^{q-1}dx$, B for the β-function $B(p,q) = \beta(p,q;1)$, and ${}_nC_{rs} = n!/[(r-1)!(s-r-1)!(n-s)!]$, the weights $a_{i,r}$, $b_{ij,rs}$ and $c_{i,rs}$ are given by

$$
\begin{aligned}
a_{i,r} &= \frac{1}{B(r,n-r+1)}\Big[\beta\Big(r,n-r+1;\frac{i}{n}\Big)-\beta\Big(r,n-r+1;\frac{i-1}{n}\Big)\Big],\\
b_{ij,rs} &= {}_nC_{rs}\sum_{k=0}^{s-r-1}\binom{s-r-1}{k}\frac{(-1)^{s-r-1-k}}{s-k-1}\Big[\Big(\frac{i}{n}\Big)^{s-k-1}-\Big(\frac{i-1}{n}\Big)^{s-k-1}\Big]\\
&\quad\times\Big[\beta\Big(k+1,n-s+1;\frac{j}{n}\Big)-\beta\Big(k+1,n-s+1;\frac{j-1}{n}\Big)\Big],\\
c_{i,rs} &= {}_nC_{rs}\sum_{k=0}^{s-r-1}\binom{s-r-1}{k}\frac{(-1)^{s-r-1-k}}{s-k-1}\Big[\beta\Big(s,n-s+1;\frac{i}{n}\Big)-\beta\Big(s,n-s+1;\frac{i-1}{n}\Big)\\
&\quad-\Big(\frac{i-1}{n}\Big)^{s-k-1}\Big(\beta\Big(k+1,n-s+1;\frac{i}{n}\Big)-\beta\Big(k+1,n-s+1;\frac{i-1}{n}\Big)\Big)\Big].
\end{aligned}
$$

3 Comparison of Estimators

To evaluate the quantile estimator $L_n(p)$ in (1), we study the relative MSE

$$\Delta\text{MSE}(L_n(p)) = \frac{\text{MSE}(\hat{Q}_n(p))}{\text{MSE}(L_n(p))} = \frac{\mathbf{E}\{(\hat{Q}_n(p)-Q(p))^2\}}{\mathbf{E}\{(L_n(p)-Q(p))^2\}}$$

compared with the sample quantile $\hat{Q}_n(p)$. Of course, if $\Delta\text{MSE}(L_n(p))$ is larger than 1, then $L_n(p)$ is better than $\hat{Q}_n(p)$. To calculate the relative MSE in turn, following e.g., [2] and [3], we use Monte Carlo simulations.

We study samples $X_1,\ldots,X_n$ from a standard normal distribution, a Student t distribution with 4 degrees of freedom, a standard lognormal distribution, and a standard exponential distribution, with sample sizes $n=25$, 50 and 100, and for the quantiles $p=0.05$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 0.95. For each cases, we use 10 000 Monte Carlo simulations, as initial studies indicated that this was required to get sufficiently narrow confidence intervals, ± 0.02 for $\Delta\text{MSE}(L_n(p))$.

In each Monte Carlo simulation, we find the weights $w_{\lfloor np\rfloor}$, $w_{\lfloor np\rfloor+1}$ and $w_{\lfloor np\rfloor+2}$ in (1) that maximize $\Delta\text{MSE}_{\text{ Boot}}(L_n(p))$, see Section 2, using the routine `fminsearch` of Matlab, with initial values $w_{\lfloor np\rfloor+1}=1$ and $w_{\lfloor np\rfloor}=w_{\lfloor np\rfloor+2}=0$, i.e., $L_n(p)=\hat{Q}_n(p)$.

4 Numerical Results and Conclusions

The results of our simulations are displayed in Figures 1-4 in Appendix 4. There we see that the estimator $L_n(p)$ almost always performs better than the sample quantile. However, the typical gain is not more than a 5% increase or so, of the relative MSE.

An study of the weights $w_{\lfloor np\rfloor}$, $w_{\lfloor np\rfloor+1}$ and $w_{\lfloor np\rfloor+2}$ obtained showed that, on the average, they were about equal in size, but that their variance were quite high.

Acknowledgments

The author is thankful to J.M.P. Albin for suggesting the problem.

References

[1] Hutson, A.D. and Ernst, M.D. (2000). The exact bootstrap mean and variance of an L-estimator. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 89-94.

[2] Sheather, S.J. and Marron, J.S. (1990). Kernal quantile estimators. *J. Amer. Statist. Assoc.* **85** 410-416.

[3] Yang, S.S. (1985). A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc.* **80** 1004-1011.

Erik Brodin
School of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
ebrodin@math.chalmers.se
http://www.math.chalmers.se/~brodin

A Figures

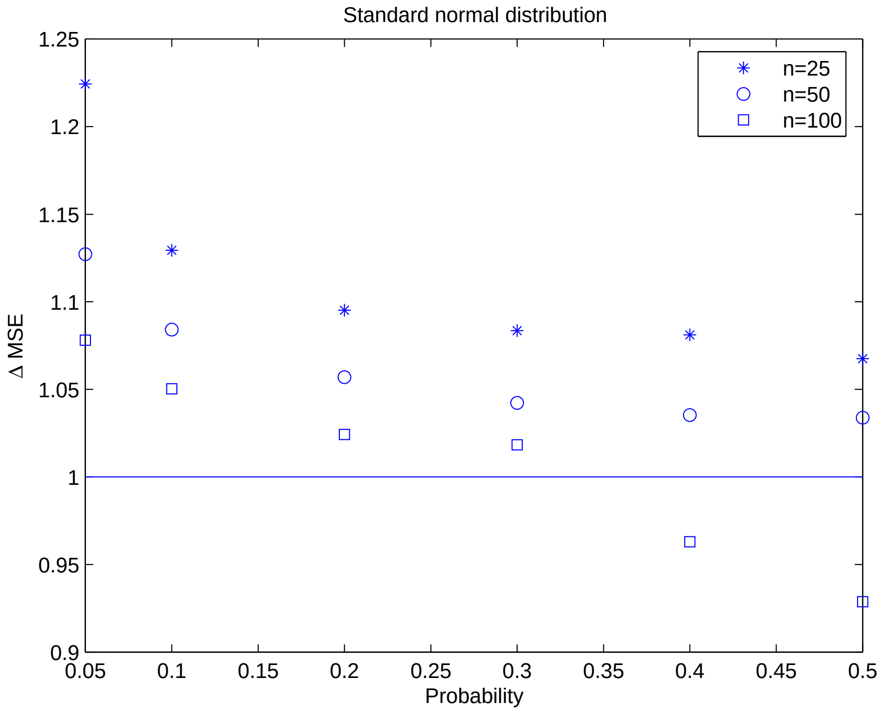

Figure 1: $\Delta\text{MSE}(L_n(p))$ for standard normal distribution.

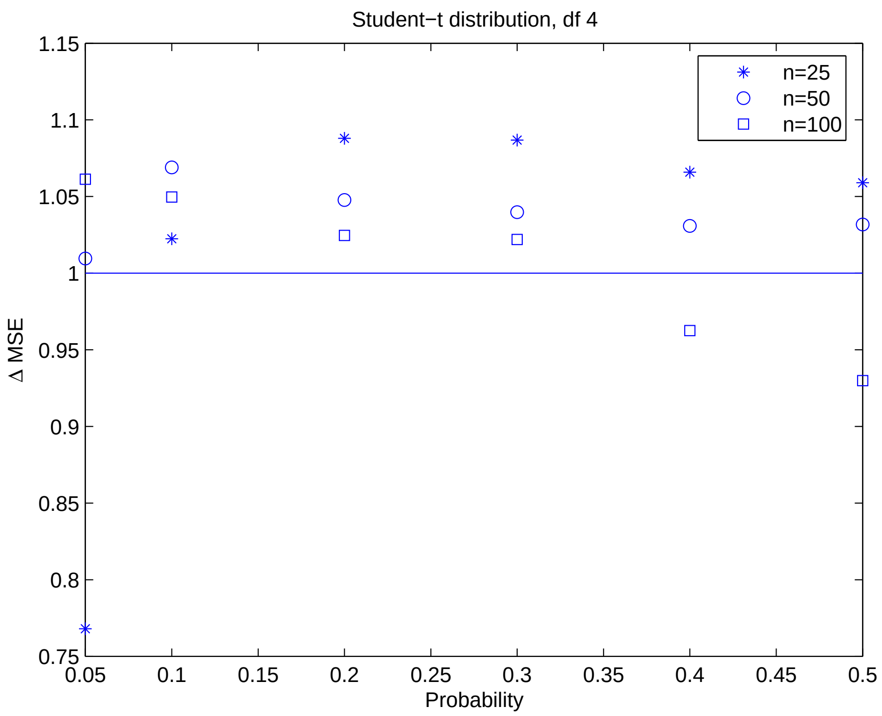

Figure 2: $\Delta\text{MSE}(L_n(p))$ for Student t distribution.

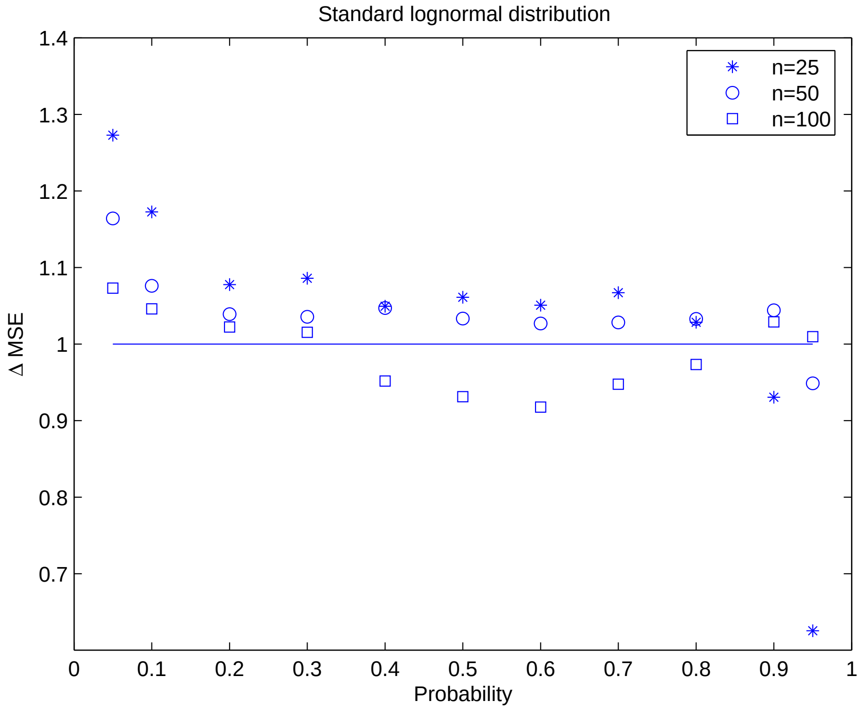

Figure 3: $\Delta\mathrm{MSE}(L_n(p))$ for standard lognormal distribution.

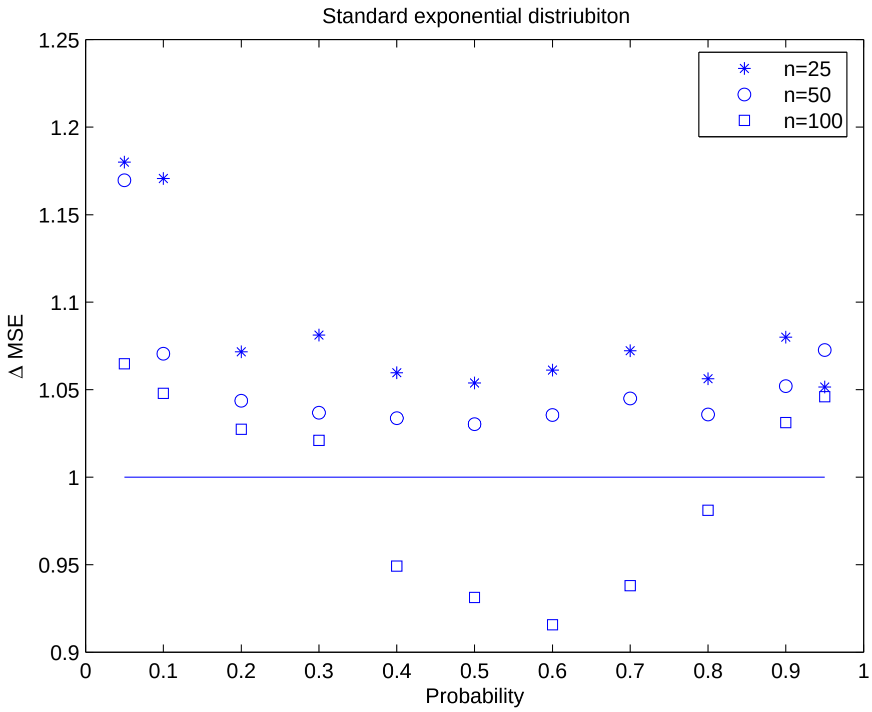

Figure 4: $\Delta\mathrm{MSE}(L_n(p))$ for standard exponential distribution.

PAPER III

Estimating Extreme Quantiles using Cross Validation and Second Order Regular Variation

Erik Brodin

Dept. of Math., Chalmers Univ. of Tech., SE-412 96 Gothenburg, Sweden

ebrodin@math.chalmers.se

Abstract

We propose a new estimator for extreme quantiles of heavy tailed distributions, based on first and second order regular variation together with cross validation. A simulation study shows that the first order regular variation estimator is more stable than the Weissman quantile estimator together with the Hill estimator of the extreme value parameter. However, we have a bias. For the second order approach the bias is small. We also test our methods on insurance loss data.

Keywords: Cross validation; Extreme value theory; Hill estimator; Quantile estimator; Regular variation; Second order regular variation; Weissman estimator.

1 Introduction

A probability distribution function F is said to belong to a domain of attraction of an extreme value distribution, if there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$(F(a_n(\cdot) - b_n))^n \stackrel{\mathrm{d}}{\to} H \quad \text{as } n \to \infty, \tag{1}$$

for some non-degenerate probability distribution function H, where $\stackrel{\mathrm{d}}{\to}$ denotes convergence in law. The possible forms of H are given by

$$H_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\Big\{-\Big(1+\gamma\,\dfrac{x-\mu}{\sigma}\Big)^{-1/\gamma}\Big\} & \text{for } \gamma\,\dfrac{x-\mu}{\sigma} > -1 \quad \text{if } \gamma \neq 0, \\ \exp\Big\{-\exp\Big[-\dfrac{x-\mu}{\sigma}\Big]\Big\} & \text{for } x \in \mathbb{R} \qquad\qquad \text{if } \gamma = 0, \end{cases} \tag{2}$$

which we denote $F \in D(H_{\gamma,\mu,\sigma})$. Here $\gamma \in \mathbb{R}$ is the extreme value parameter, $\mu \in \mathbb{R}$ a location parameter, and $\sigma > 0$ a scale parameter. When $\gamma > 0$, F is called heavy tailed.

In practice, it is important to find the distribution of a rare outcome, given that it really occurs, in order to estimate extreme quantiles. As this involves extrapolation beyond the range of a finite sample, one has to assume a model. For reasons that can be seen in Section 2, a natural model is an extreme value distribution $H_{\gamma,\mu,\sigma}$.

To fit an extreme value distribution to a sample, one selects a threshold and assumes that the part of the data that exceeds the threshold has distribution $H_{\gamma,\mu,\sigma}$. However, selecting a suitable threshold is a non-trivial problem.

We use a cross validation scheme to estimate extreme quantiles, which makes the selection of threshold less crucial. To control bias, we consider a second order setting for the convergence (1) when $\gamma > 0$, which leads us to a new estimator of extreme quantiles.

The disposition of the article is as follows: Section 2 surveys estimation of extreme quantiles. In Section 3 we discuss estimation of the extreme value parameter γ. Sections 4 and 5 introduce second order theory and a new estimator of extreme quantiles for $\gamma > 0$. In Section 6 we discuss extensions to $\gamma \in \mathbb{R}$. Section 7 describes the cross validation scheme. In Sections 8 and 9 we evaluate the new estimator by simulations, and test the estimator on insurance loss data. Section 10 concludes our article.

2 Quantile Estimation for Heavy Tails

We will here shortly present the most common methods to estimate extreme quantiles for heavy tails. See Embrechts et al. (1997) and Coles (2001) on more information. It should be noted that all these methods require selection of thresholds, in different ways.

For the quantile $x_p = H^{-1}_{\gamma,\mu,\sigma}(1-p)$ in (2), we have

$$x_p = \mu - \frac{\sigma}{\gamma}\left(1 - (-\log(1-p))^{-\gamma}\right) \quad \text{for } \gamma > 0.$$

So, given estimates $\hat{\gamma}$, $\hat{\mu}$ and $\hat{\sigma}$ of γ, μ and σ, we can estimate x_p by

$$x_p^{\mathrm{EV}} = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}\left(1 - (-\log(1-p))^{-\hat{\gamma}}\right).$$

Here γ, μ and σ can be estimated by maximum likelihood methods, under suitable independence assumptions between data blocks, for extreme observations among data.

The peaks over threshold method uses the conditional distribution $F_u(x) = F(x-u)/(1-F(u))$ given exceedance of a threshold u: Pickands (1975) showed that

$$F \in D(H_{\gamma,\mu,\sigma}) \quad \Leftrightarrow \quad \lim_{u \uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\gamma,0,\sigma(u)}(x)| = 0.$$

Here $\sigma(u) > 0$ is a suitable function and $G_{\gamma,\mu,\sigma}$ the generalized Pareto distribution

$$G_{\gamma,\mu,\sigma}(x) = \begin{cases} 1 - \left[0 \vee \left(1 + \gamma \dfrac{x-\mu}{\sigma}\right)\right]^{-1/\gamma} & \text{if } \gamma \neq 0, \\ 1 - \exp\left\{-\dfrac{x-\mu}{\sigma}\right\} & \text{if } \gamma = 0, \end{cases}$$

while $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ is the right end point of F. When $\gamma \neq 0$, this gives

$$1 - F(x) \approx (1 - F(u)) \left(1 + \gamma \frac{x-u}{\sigma}\right)^{-1/\gamma} \quad \text{for } x > u, \text{ for } u \text{ sufficiently large.}$$

Here $1 - F(u)$ can be estimated by the rate N_u/n of exceedences of u by a sample of independent random variables $X_1, \ldots, X_n$ with distribution function F. So, given estimators $\hat{\gamma}$ and $\hat{\sigma}$ of γ and σ (see e.g., Smith, 1987), we can estimate $x_p = F^{-1}(1-p)$ by

$$\hat{x}_p^{\text{POT}} = u + \hat{\sigma} \frac{(N_u/(np))^{\hat{\gamma}} - 1}{\hat{\gamma}}.$$

For $\gamma > 0$, we have $F \in D(H_{\gamma,0,0})$ if and only if

$$F^{-1}(1 - 1/x)x^{-\gamma} = l(x) \quad \text{where} \quad \lim_{x\to\infty} \frac{l(tx)}{l(x)} = 1 \quad \text{for } t > 0. \tag{3}$$

This means that $F^{-1}(1 - 1/(\cdot))$ is a regularly varying function with index γ, and that l is a slowly varying function. Now, picking a sufficiently large order statistic $X_{(n-k)}$, we may derive the following estimator of $x_p = F^{-1}(1-p)$ proposed by Weissman (1978)

$$\hat{x}_p^{\text{reg}} = X_{(n-k)} \left(\frac{k+1}{(n+1)p}\right)^{\hat{\gamma}}. \tag{4}$$

3 Estimation of the Extreme Value Parameter

Although it is arguable more important to estimate quantiles in practice, it is estimation of γ that has recived the most attention in the literature.

Using that F^{-1} is regularly varying for a heavy tailed distribution, we can write

$$\frac{F^{-1}(1 - 1/(tx))}{F^{-1}(1 - 1/x)} \approx t^{\gamma} \quad \text{for } x > 0 \text{ sufficiently large.}$$

Taking logarithms, we get the following information to estimate γ:

$$\log\left(\frac{X_{(n-k+i)}}{X_{n-k}}\right) \approx \gamma \log\left(\frac{k+1}{k+1-i}\right) \quad \text{for } i = 1, \ldots, k.$$

See Kratz and Resnick (1996) and Schultze and Steinebach (1996) on more on this.

The most common estimator of γ is the Hill estimator (see Embrechts et al., 1997)

$$\hat{\gamma}_k^{\text{Hill}} = \frac{1}{k} \sum_{j=1}^{k} \left(\log(X_{(n-j+1)}) - \log(X_{(n-k)})\right) \quad \text{for } k = 1, \ldots, n-1.$$

We will denote with $\hat{x}_p^{\text{Hill}}$ a quantile estimator that uses $\hat{x}_p^{\text{reg}}$ from (4) with $\hat{\gamma} = \hat{\gamma}_k^{\text{Hill}}$.

4 Second Order Extensions

The convergence in (3) can be slow, which leads to a bias when estimating γ or the quantile x_p. Several methods have been developed to deal with this problem. For example, there is literature on optimal selection of thresholds, to get smaller bias. See e.g., Caers et al. (1998), Danielsson et al. (2001) and Drees and Kaufmann (1998).

Feuerverger and Hall (1999) proposed a refinement of the model (3), by considering the Hall class of Pareto type models (see Hall and Welsh, 1985),

$$1 - F(x) = Cx^{-1/\gamma}(1 + Dx^{\rho/\gamma} + o(x^{\rho/\gamma})),$$

with parameters $C > 0$, $D \in \mathbb{R}$, and $\rho < 0$, which includes most common heavy tailed distributions. For this model they developed estimators of γ, based on the regression

$$i\left(\log(X_{(n-i+1)}) - \log(X_{(n-i)})\right) \approx Z_i \gamma \exp\{D(n/i)^\rho\} \quad \text{for } 1 \le i \le k,$$

where Z_i are independent standard exponential random variables and k a threshold.

Independently, Beirlant et al. (1999) used the assumption on the function l in (3), that there exist a constant $\rho < 0$ and a function $b > 0$ with $\lim_{x\to\infty} b(x) = 0$, such that

$$\lim_{x\to\infty} \frac{1}{b(x)} \log\left(\frac{l(tx)}{l(x)}\right) = h_\rho(\lambda) \quad \text{for } t \ge 1, \tag{5}$$

where $h_\rho(t) = (t^\rho - 1)/\rho$, which again includes most common heavy tailed distributions. With this assumption they derived the regression

$$i\left(\log(X_{(n-i+1)}) - \log(X_{(n-i)})\right) = Z_i\left(\gamma + B((k+1)/i)^\rho\right) \quad \text{for } 1 \le i \le k,$$

where B is a constant that replaces the function b.

Matthys et al. (2004) used (5) to derive the quantile estimator

$$\hat{x}_p^{\text{Matt}} = X_{(n-k)}\left(\frac{k+1}{(n+1)p}\right)^{\hat{\gamma}} \exp\left\{\hat{B}\frac{1 - ((n+1)p/(k+1))^{-\hat{\rho}}}{-\hat{\rho}}\right\}.$$

5 A New Estimator

Although (5) is based on second order regular variation, it is not identical to the definition of that concept of de Haan and Statdmüller (1996), that we will use:

$$\lim_{x\to\infty} \frac{1}{A(x)}\left(\frac{F^{-1}(1 - 1/(xt))}{F^{-1}(1 - 1/x)} - t^\gamma\right) = t^\gamma h_\rho(t) \quad \text{for } t > 0. \tag{6}$$

Here the second order parameter $\rho \le 0$ dictates the rate of convergence in (3), while A

is a normalizing function that is regularly varying with index ρ and $\lim_{x\to\infty} A(x) = 0$. Further, $h_\rho(t) = (t^\rho - 1)/\rho$ for $\rho < 0$ and $h_0(t) = log(t)$.

Taking $A(x) = cx^\rho$ for a constant $c \in \mathbb{R}$, in (6), we get the following estimator of x_p:

$$\hat{x}_p^{\text{2-reg}} = X_{(n-k)} \left(\frac{k+1}{(n+1)p}\right)^{\hat{\gamma}} \left[1 + \hat{c}\left(\frac{n+1}{k+1}\right)^{\hat{\rho}} h_{\hat{\rho}}\left(\frac{k+1}{(n+1)p}\right)\right]. \tag{7}$$

This estimator coincides with that obtained from using a first order expansion of the exponential function in $\hat{x}_p^{\text{Matt}}$, together with $\hat{B} = \hat{c}((n+1)/(k+1))^{\hat{\rho}}$. And so $\hat{x}_p^{\text{2-reg}}$ takes into account that the function A is regularly varying, which $\hat{x}_p^{\text{Matt}}$ does not.

Notice that $\hat{x}_p^{\text{2-reg}}$ is based on the the assumption on the function l in (3), that

$$\frac{l(tx)}{l(x)} \approx 1 + cx^\rho h_\rho(t). \tag{8}$$

This is called slow variation with a reminder SR2 by Bingham et al. (1987), p. 185.

By (3) and (8), together with straightforward calculations and the fact that $X_{(n-k)} \stackrel{\text{d}}{=} F^{-1}(1 - 1/U_{(k+1)}^{-1})$, where $\stackrel{\text{d}}{=}$ is equality in distribution and $U_{(1)} \leq \ldots \leq U_{(n)}$ are order statistics of a uniformly distributed sample over $(0,1)$, we get

$$\log\left(\frac{X_{(n-j+1)}}{X_{(n-j)}}\right) \stackrel{\text{d}}{=} \log\left[\left(\frac{U_{(j+1)}}{U_{(j)}}\right)^\gamma \left(1 + c\left(U_{(j+1)}^{-1}\right)^\rho h_\rho\left(\frac{U_{(j+1)}}{U_{(j)}}\right)\right)\right] \quad \text{for } j = 1, \ldots, k.$$

By a result of S. Malmquist (see e.g., David, 1981), $Z_j = U_{(j)}/U_{(j+1)}$ are independent and uniformly distributed over $(0,1)$. Replacing $U_{(j+1)}^{-1}$ with $E\{U_{(j+1)}^{-1}\} \approx 1/E\{U_{(j+1)}\}$ $= (n+1)/(j+1)$, we get an approximation useful for maximum likelihood estimation:

$$\log\left(\frac{X_{(n-j+1)}}{X_{(n-j)}}\right) \stackrel{d}{\approx} \log\left[Z_j^{-\gamma/j}\left(1 + c\left(\frac{n+1}{j+1}\right)^\rho \frac{Z_j^{-\rho/j} - 1}{\rho}\right)\right] \quad \text{for } j = 1, \ldots, k.$$

We are primarily interested in the bias for finite samples rather than asymptotics. However, under suitable conditions, $\hat{x}_p^{\text{reg}}$ and $\hat{x}_p^{\text{2-reg}}$ have the same asymptotic properties:

Proposition 5.1. *Consider (7) with $\rho < 0$. Let $\hat{\rho}$ be a consistent in probability estimator of ρ and $\hat{c}$ a bounded in probability estimator of c. For a sequence $a(k) \to \infty$ as $k \to \infty$ such that $(k+1)/(n+1) \to 0$, $(k+1)/((n+1)p) \to \infty$ and $a_k((n+1)/(k+1))^{\rho+\epsilon} \to 0$ for some $\epsilon > 0$, we have*

$$a_k(\hat{x}_p^{reg} - x_p) \stackrel{d}{\to} \mathrm{N}(0,\sigma^2) \quad \Leftrightarrow \quad a_k(\hat{x}_p^{2\text{-}reg} - x_p) \stackrel{d}{\to} N(0,\sigma^2).$$

Proof. We have

$$a_k(\hat{x}_p^{\text{2-reg}} - x_p)$$
$$= a_k(\hat{x}_p^{\text{reg}} - x_p) + a_k(\hat{x}_p^{\text{reg}} - x_p)\hat{c}\Big(\frac{n+1}{k+1}\Big)^{\hat{\rho}} h_{\hat{\rho}}\Big(\frac{k+1}{(n+1)p}\Big) + a_k x_p \hat{c}\Big(\frac{n+1}{k+1}\Big)^{\hat{\rho}} h_{\hat{\rho}}\Big(\frac{k+1}{(n+1)p}\Big).$$

Here the second and third terms on the right-hand side go to zero in probability, as a more or less immediate consequence of the hypothesis. □

6 More General Tails

Here we discuss methology to estimate quantiles for $\gamma \in \mathbb{R}$.

For $\gamma \in \mathbb{R}$, we have $F \in D(H_{\gamma,0,1})$ if and only if (see de Haan, 1984)

$$\lim_{x\to\infty} \frac{F^{-1}(1-1/(xt)) - F^{-1}(1-/x)}{a(x)} = \frac{t^{\gamma}-1}{\gamma} \quad \text{for } t>0. \tag{9}$$

Here $a > 0$ is a normalizing function, and for $\gamma = 0$ the right hand side should be interpreted as $\log(t)$. This gives the following estimator of x_p:

$$\tilde{x}_p^{\text{reg}} = X_{(n-k)} + \hat{a}((k+1)/(n+1))\frac{((k+1)/(n+1)p)^{\hat{\gamma}}-1}{\hat{\gamma}}$$

Drees (2003) suggests the the following estimator of a:

$$\hat{a}(k/n) = \frac{\hat{\gamma}}{2^{\hat{\gamma}}-1}(X_{n-\lfloor k/2\rfloor} - X_{(n-k)}),$$

where $\lfloor x\rfloor$ is the integer part of x. This estimator could be made more flexible by using $X_{n-\lfloor k/m\rfloor}$ instead of $X_{n-\lfloor k/2\rfloor}$, where m is selected as a part of the cross validation.

As for (3), the convergence in (9) can be slow. To handle this, we can assume that

$$\lim_{x\to\infty}\frac{1}{A(x)}\left(\frac{F^{-1}(1-1/xt) - F^{-1}(1-1/x)}{a(x)} - \frac{t^{\gamma}-1}{\gamma}\right) = \frac{1}{\rho}\left(\frac{t^{(\gamma+\rho)}-1}{\gamma+\rho} - \frac{t^{\gamma}-1}{\gamma}\right)$$

for $t > 0$: Here A is regularly varying at infinity with index $\rho \le 0$ and $\lim_{x\to\infty} A(x) = 0$. If we reason as when constructing $\hat{x}_p^{\text{2-reg}}$, this gives a second order estimator $\tilde{x}_p^{\text{2-reg}}$ of x_p.

7 Estimating Quantiles by Cross Validation

Here we propose a cross validation method for parameter estimates.

Cross validation requires a score function to minimize. Naturally, the selection of this function is both important and difficult. We will minimize the distance between estimated extreme quantiles and order statistics. This interprets as minimize prediction of low probability events, with empirical quantiles as benchmark.

For a sample of size n, and using K order statistics, we minimize the score function

$$CV = \sum_{k=0}^{K-1} |X_{(n-k)} - \hat{x}_{(k+1)/(n+1)}|. \tag{10}$$

Here we use $\hat{x}_p^{\text{reg}}$ or $\hat{x}_p^{\text{2-reg}}$ as estimators $\hat{x}_p$ of x_p. The resulting estimators are denoted $\hat{x}_p^{\text{cross}}$ and $\hat{x}_p^{\text{2-reg}}$, respectively. Although one has to select how many tail data K to use, that number typically affects results much less than do thresholds.

Of course, one could consider a weighted sum in (10). Also notice that (10) in a way is a L^1 regression.

When using the cross validation, we minimize CV both with respect to the parameters to be estimated, and with respect to threshold selections for $\hat{x}_p^{\text{reg}}$ and $\hat{x}_p^{\text{2-reg}}$

8 Testing the Estimators $\hat{x}_p^{\text{cross}}$ and $\hat{x}_p^{\text{2-reg}}$

We use 100 simulations to compare $\hat{x}_p^{\text{cross}}$ and $\hat{x}_p^{\text{2-reg}}$ with $\hat{x}_p^{\text{Hill}}$ for $n = 500$ data. We involve 30-120 order statistics in CV and $\hat{x}_p^{\text{Hill}}$. The following distributions are used:

- Burr distribution: The Burr(β, τ, λ) distribution function is given by

$$F(x) = 1 - \Big(\frac{\beta}{\beta + x^\tau}\Big)^\lambda \quad \text{for } x \geq 0.$$

  We use Burr$(1, 2, 0.5)$, Burr$(1, 1, 1)$ and Burr$(1, 0.5, 2)$ to model really heavy tails ($\gamma = 1$ and $\rho = -0.5, -1, -2$), and Burr$(1, \frac{20}{3}, 0.5)$, Burr$(1, \frac{10}{3}, 1)$ and Burr$(1, \frac{5}{3}, 2)$ to model moderately heavy tails ($\gamma = 0.3$ and $\rho = -0.5, -1, -2$).

- Student t distribution with ν degrees of freedom: We use $\nu = 2$, 4 and 8 ($\gamma = 0.5$, 0.25 and 0.125 and $\rho = -1$, -0.5 and -0.25).

We used the `fminsearch` command of Matlab for our optimizations, with starting values $\gamma = 0.5$, $\rho = -1$ and $c = 0$, and with starting thresholds the first positive value for the Student t distribution, and $X_{(1)}$ for the Burr distribution.

We only present results of our simulations for the 0.999-quantile. Figures that display the results are collected in Appendix 10. They show the median, as indicator of bias, and the 25% and 75% quartiles, as indictors of variation (variance). Results for the 0.9999-quantile are similar.

For $\hat{x}_p^{\text{cross}}$, Figures 5-7 show that the cross validation approach works better than $\hat{x}_p^{\text{Hill}}$ for small values of γ, but poorer for γ large, as the latter gives a large variance. This is not surprising because the cross validation function should be more stable the more moments that exist. However, $\hat{x}_p^{\text{cross}}$ consistently has a smaller bias than $\hat{x}_p^{\text{Hill}}$.

For $\hat{x}_p^{\text{2-reg}}$, Figures 2-10 show that the bias is much smaller than for $\hat{x}_p^{\text{Hill}}$, but the variance slightly greater, except for really heavy tails where it is much greater.

9 Application to Insurance Loss Data

We test the cross validation method on Danish fire loss data. The data set contains of 2157 losses over one million Danish Krone (DKK) from 1980 to 1990. McNeil (1997) studied this data and found that estimated extreme quantiles, like 0.999 and 0.9999, depended heavily on the number of order statistics involved in the estimation.

In Figures 11 and 12 we can see the estimated 0.999 and 0.9999 quantiles. Notice the stable behavior of the cross valuation method compared to that of $\hat{x}_p^{\mathrm{Hill}}$. Further, $\hat{x}_p^{\mathrm{cross}}$ and $\hat{x}_p^{\text{2-reg}}$ are very similar when more than 200 order statistics are used.

10 Conclusions

We have introduced cross validation for estimation of extreme quantiles, as well as a new estimator based on second order regular variation. The cross validation method works well for moderately heavy tails, and is stable with respect to the number of order statistics involved in the estimation. The new estimator displays a small bias compared to the estimators based on ordinary regular variation.

Acknowledgments

The author would like to thank Patrik Albion for improving the manuscript from an early draft. The author also would like to thank Holger Rootzén.

References

Beirlant, J., Dierckx, G. Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2** 177-200.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

Caers, J., Beirlant, J. and Vynckier, P. (1998). Bootstrap confidence intervals for tail indices. *Comput. Statist. Data Anal.* **26** 259-277.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer-Verlag, Berlin.

Danielsson, J., de Haan, L., Peng, L. and Vries C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.*

76 226-248.

David, H.A. (1981). *Order Statistics (2nd Ed.).* John Wiley, New York.

Drees, H. (2003). Extreme quantile estimation for dependent data with application to finance. *Bernoulli* **9** 617-657.

Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.* **75** 149-172.

Embrechts, P.A.L., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.,* **27** 760-781.

de Haan, L. (1984). Slow variation and characterization of domains of attraction. In: J. Tiago de Oliveira (Ed.) *Statistical Extremes and Applications,* Reidel, Dordrecht, 31-48.

de Haan, L. and Stadtmüller, U. (1996). Generalized regular variatioon of second order. *J. Austral. Math. Soc.* **61** 381-395.

Hall, P. and Welsh A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 331-341.

Kratz, M. and Resnick S. (1996). The qq-estimator and heavy tails. *Commun. Statist. Stochastic Models* **12** 699-724.

McNeil, A.J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bull.* **27** 117-137.

Matthys, G., Delafosse, E., Guillou, A. and Beirlant J. (2004). Estimating catatrophic quantile levels for heavy-tailed distributions. *Insurance Nath. Econom.* **34** 517-537.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119-131.

Schultze, J. and Steinebach J. (1996). On least squares estimates of an exponential tail coefficent. *Statist. Decisions* **14** 353-372.

Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174-1207.

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

A Figures

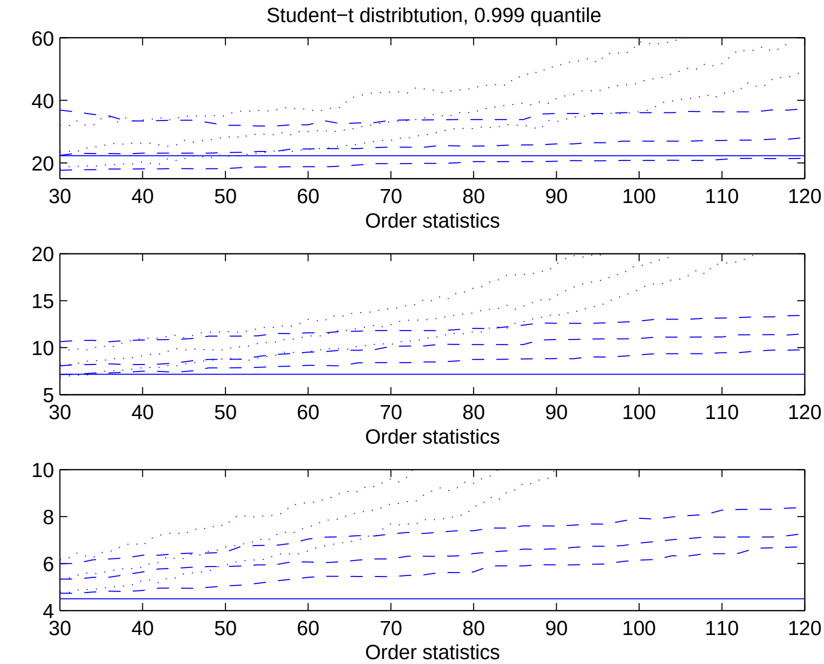

Figure 5: True quantile (straight line), and median and quartiles of $\hat{x}_p^{\text{Hill}}$ (dotted lines) and $\hat{x}_p^{\text{cross}}$ (dashed lines). From the top 2, 4 and 8 degrees of freedom.

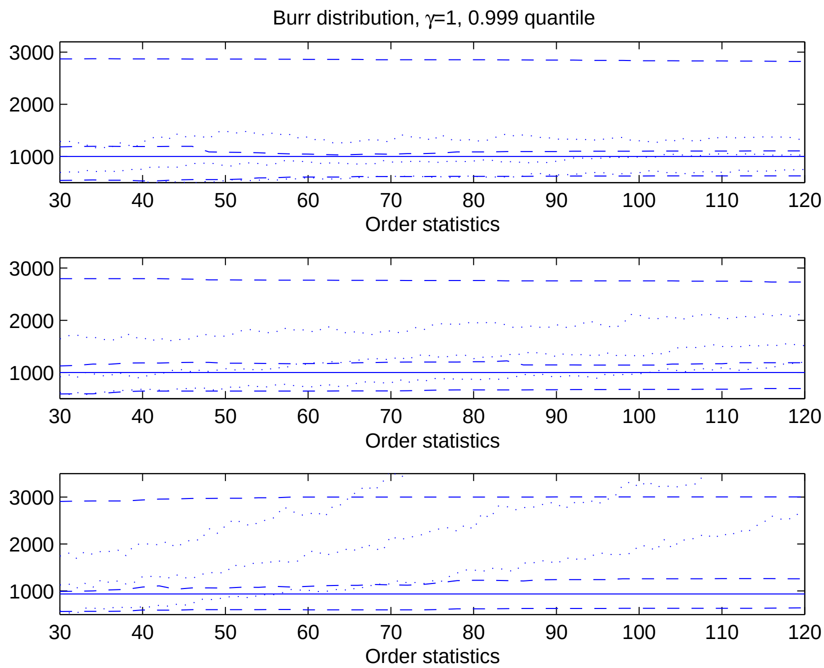

Figure 6: True quantile (straight line), and median and quartiles of $\hat{x}_p^{\text{Hill}}$ (dotted lines) and $\hat{x}_p^{\text{cross}}$ (dashed lines). From the top Burr$(1, 2, 0.5)$, Burr$(1, 1, 1)$ and Burr$(1, 0.5, 2)$.

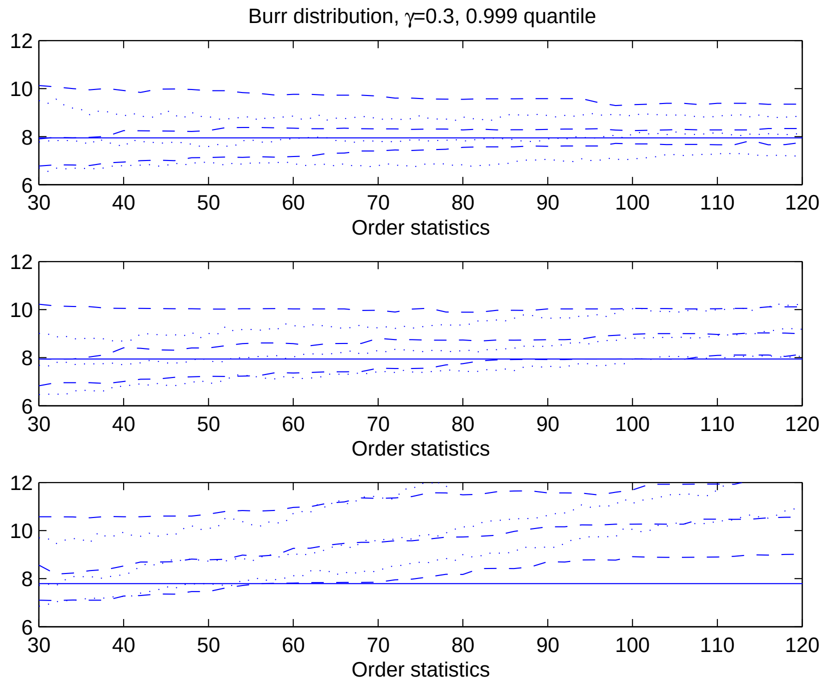

Figure 7: True quantile (straight line), and median and quartiles of $\hat{x}_p^{\text{Hill}}$ (dotted lines) and $\hat{x}_p^{\text{cross}}$ (dashed lines). From the top Burr$(1, \frac{20}{3}, 0.5)$, Burr$(1, \frac{10}{3}, 1)$ and Burr$(1, \frac{5}{3}, 2)$.

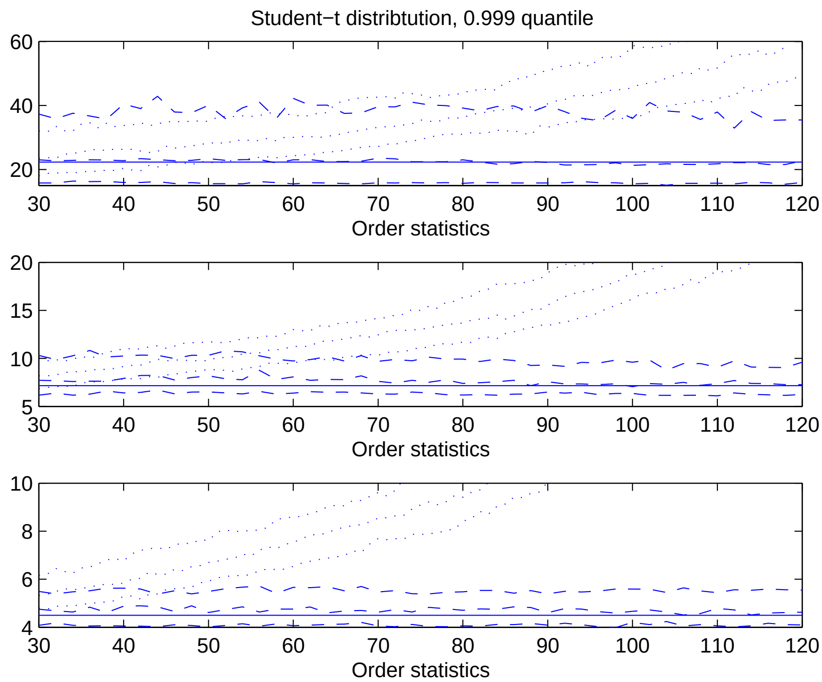

Figure 8: True quantile (straight line), and median and quartiles of $\hat{x}_p^{\text{Hill}}$ (dotted lines) and $\hat{x}_p^{\text{2-reg}}$ (dashed lines). From the top 2, 4 and 8 degrees of freedom.

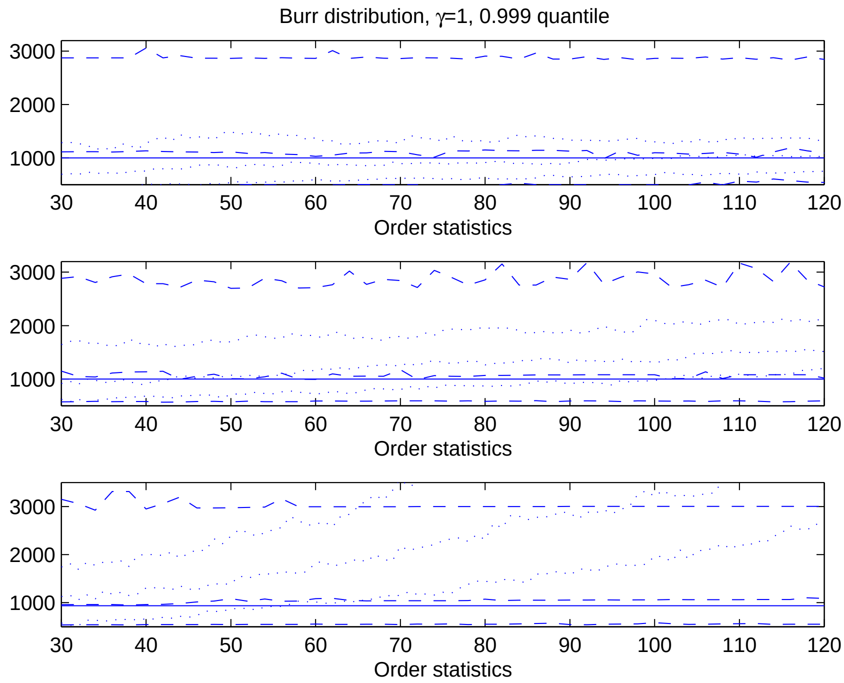

Figure 9: True quantile (straight line), and median and quartiles of $\hat{x}_p^{\text{Hill}}$ (dotted lines) and $\hat{x}_p^{\text{2-reg}}$ (dashed lines). From the top Burr$(1, 2, 0.5)$, Burr$(1, 1, 1)$ and Burr$(1, 0.5, 2)$.

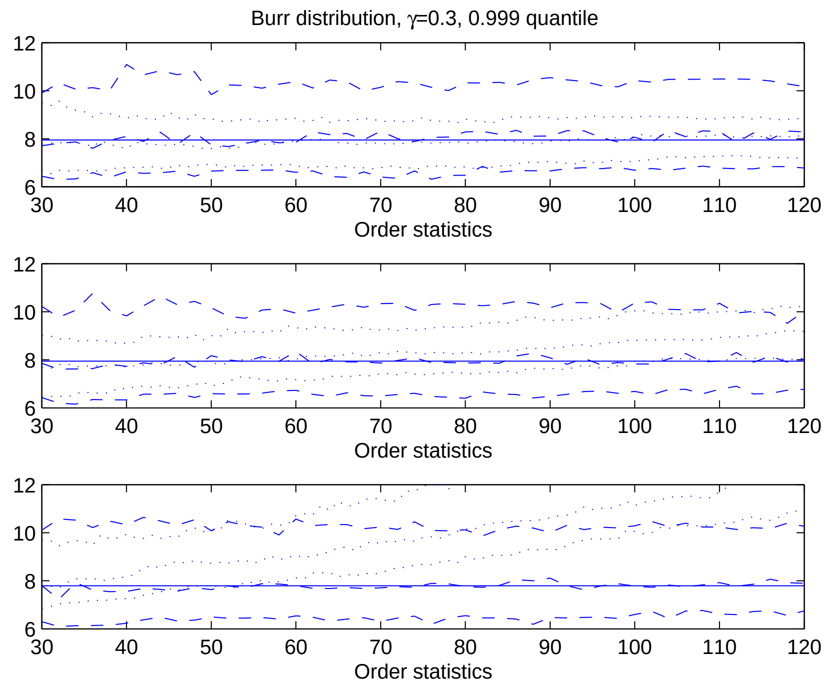

Figure 10: True quantile (straight line), and median and quartiles of $\hat{x}_p^{\text{Hill}}$ (dotted lines) and $\hat{x}_p^{\text{2-reg}}$ (dashed lines). From the top Burr$(1, \frac{20}{3}, 0.5)$, Burr$(1, \frac{10}{3}, 1)$ and Burr$(1, \frac{5}{3}, 2)$.

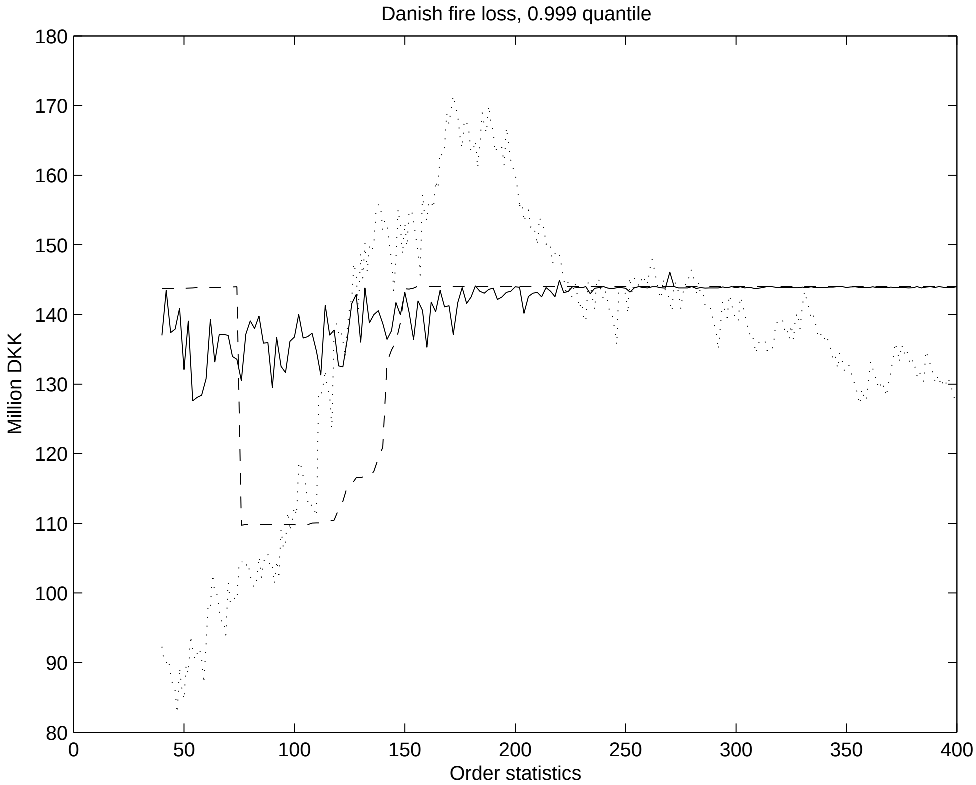

Figure 11: $\hat{x}_p^{\mathrm{Hill}}$ (dotted line), $\hat{x}_p^{\mathrm{cross}}$ (dashed line) and $\hat{x}_p^{\text{2-reg}}$ (full line).

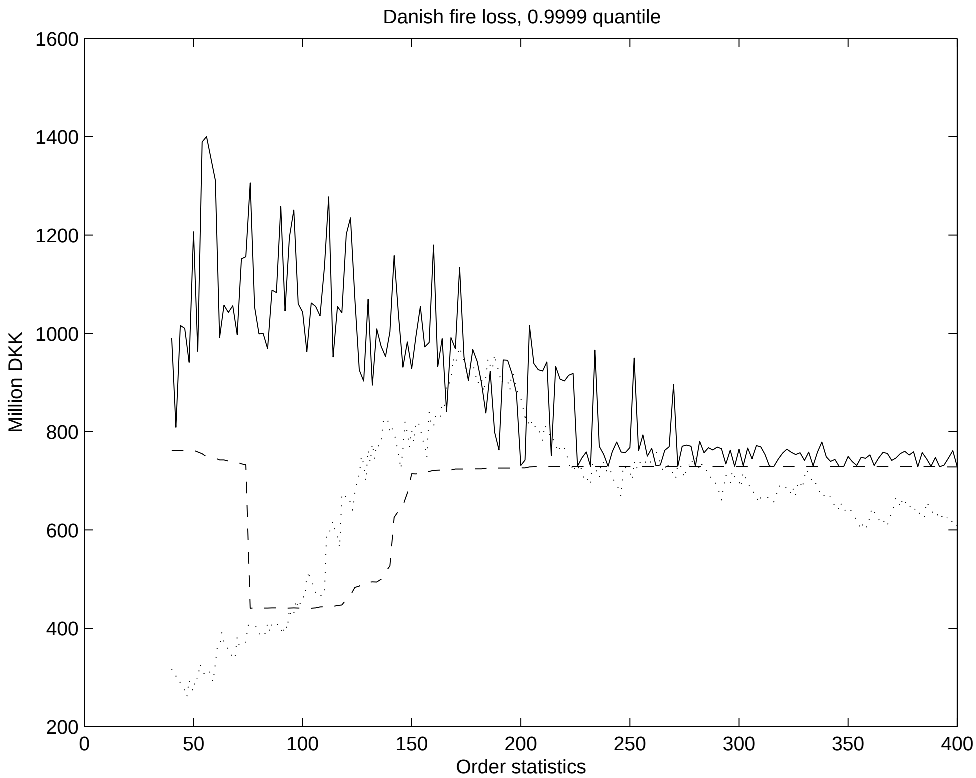

Figure 12: $\hat{x}_p^{\mathrm{Hill}}$ (dotted line), $\hat{x}_p^{\mathrm{cross}}$ (dashed line) and $\hat{x}_p^{\text{2-reg}}$ (full line).

PAPER IV

A Note on Quantile Estimation using Perturbed Generalized Pareto Distributions[1]

Introduction and Motivation

In literature, the following methods are used to estimate high quantiles:

1. *Non-parametric method:* This method uses linear combinations of order statistics.
2. *Distribution fitting:* This method fits a distribution to the whole data set.
3. *Extreme value theory:* This methods fits an extreme value distribution to the uppermost order statistics.

However, these methods have drawbacks: Non-parametric methods have high variance in tails, distribution fittings use a whole sample which is unsuitable when interested in tails, and with extreme value theory it is difficult to identify where the tail begins.

Given a probability distribution F that belongs to a domain of attractions of extremes (see e.g., Embrechts et al., 1997), and writing $\bar{F} = 1 - F$, we like a model for

$$\bar{F}_u(x) = \frac{\bar{F}(u+x)}{\bar{F}(u)} \quad \text{for } x > 0, \text{ for a large threshold } u. \tag{1}$$

By a result of J. Pickands III, we can use a generalized Pareto distribution (GPD), as

$$\lim_{u \uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\gamma,0,\sigma(u)}(x)| = 0, \tag{2}$$

see (Embrechts et al., 1997): Here $x_F = \sup\{x : F(x) < 1\}$ is the right end-point of F and $\sigma(u) > 0$ a suitable function, while $G_{\gamma,\mu,\sigma}$ is the GPD probability distribution

$$G_{\gamma,\mu,\sigma}(x) = \begin{cases} 1 - \left[0 \vee \left(1 + \gamma \dfrac{x-\mu}{\sigma}\right)\right]^{-1/\gamma} & \text{if } \gamma \neq 0, \\ 1 - \exp\left\{-\dfrac{x-\mu}{\sigma}\right\} & \text{if } \gamma = 0. \end{cases} \tag{3}$$

As it is difficult to find the threshold u, we will consider a model with an additional parameter ρ that dictates the speed of convergence in (2), and thus how to select u.

Second Order Theory

Second order theory is natural to model the speed of convergence in (2), see e.g., (de Haan and Stadtmüller, 1996), (Beirlant et al., 1999) and (Feuerverger and Hall, 1999).

For a heavy tailed distribution, i.e., $\gamma > 0$ in (3), $\bar{F}$ is regularly varying, so that

[1]By Erik Brodin. Adress: Dept. of Math., Chalmers Univ. of Tech., 412 96 Gothenburg, Sweden. Email: ebrodin@math.chalmers.se

$$\bar{F}(x) = x^{-1/\gamma} l(x) \quad \text{where} \quad \lim_{x\to\infty} \frac{l(tx)}{l(x)} = 1 \quad \text{for } t > 0. \tag{4}$$

Second order theory means modelling the speed of convergence for the ratio in (4).

Heavy tailed distribution. By (1) and (4), we have

$$\bar{F}_u(x) = \left(\frac{u+x}{u}\right)^{-1/\gamma} \frac{l(u(1+x/u))}{l(u)} \quad \text{for } x > 0. \tag{5}$$

The following second order assumption is due to (Bingham et al., 1987, p. 185, SR2):

$$\frac{l(\lambda x)}{l(x)} = 1 + cx^{\rho} h_{\rho}(\lambda). \tag{6}$$

Here $h_\rho(t) = (t^{-\rho} - 1)/\rho$ while $\rho < 0$ and $c \in \mathbb{R}$ are constants. Taking $\beta = cu^\rho/\rho$ and $\sigma = u\gamma$ in (5) and (6), we get the perturbed generalized Pareto distribution (PGPD)

$$\bar{F}_u(x) = \left(1 + \gamma\frac{x}{\sigma}\right)^{-1/\gamma} (1-\beta) + \beta\left(1 + \gamma\frac{x}{\sigma}\right)^{-1/\gamma + \text{sign}(\gamma)\rho} \quad \text{for } 1 + \gamma\frac{x}{\sigma} > 0. \tag{7}$$

Light tailed distribution. For a light tailed distribution, i.e., $\gamma < 0$ in (3), we have $x_F < \infty$. Further, $\bar{F}(x_F - x^{-1})$ is regular varying, so that

$$\bar{F}(x_F - x^{-1}) = x^{1/\gamma} l(x) \quad \text{where} \quad \lim_{x\to\infty} \frac{l(tx)}{l(x)} = 1 \quad \text{for } t > 0.$$

Using (6), and taking $\beta = c(x_F - u)^{-\rho}$ and $\sigma = (u - x_F)\gamma$, we get the PGPD (7) again.

Exponential tailed distribution. We do not consider $\gamma = 0$ because of so called penultimate approximation. This method is based on replacing γ with $\gamma_n \neq \gamma$, for a finite sample of size n. And so $\gamma \neq 0$ comes into play also here. For example, the light tailed case should be used for normal distributions (which have $\gamma = 0$), see (Cohen, 1982).

Estimation

We considered parameter estimation by the maximum likelihood method, by minimizing

$$\max_{k\le i\le n} \left| F(X_{(i)}) - \frac{i-0.5}{n+1} \right|,$$

where $X_{(k)}$ is the first order statistic above the threshold u, and by minimizing

$$\max_{k\le i\le n} \left| X_{(i)} - F^{-1}\left(\frac{i-0.5}{n+1}\right) \right|.$$

Estimating $\bar{F}(u)$ by the rate N_u/n of observations above the threshold u, we get the following estimator of the quantile $x_p = F^{-1}(1-p)$:

$$\hat{x}_p = u + \overline{PGPD}^{-1}_{\hat{\gamma},0,\hat{\sigma},\hat{\rho}}\left(\frac{np}{N_u}\right).$$

Conclusions

We used Monte Carlo simulation to compare GPD and PGPD, with the different estimation schemes, for sample sizes 500 from beta, Burr, normal and Student t distributions. PGPD and GPD gave different estimates of γ, but virtually the same quantile estimates.

We can conclude that it seems to be a difficult problem to make proper use of the PGPD model to estimate extreme quantiles. This is a very surprising result, to us.

Note on priority

During the work with this article, the work (Beirlant et al., 2004a) on the case $\gamma > 0$ was published, see also (Beirlant et al., 2004b), where a one-parameter extension of GPD was introduced. There are problems with this approach for $\gamma < 0$, as x_F is unknown.

At the end of the work with this article, we discovered the unpublished manuscript (Beirlant et al., 2002), which uses the PGPD approach. At large, Beirlant et al. make the same conclusions as we do, although they get a sligtly more positive experience of the PGPD model than we do. With this note of priority, we find it well motivated to publish our article, but in a shorter format than originally intended.

References

Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2** 177-200.

Beirlant, J., Joossens, E. and Segers, J. (2002). Modelling excesses over high tresholds by perturbed generalized Pareto distributions. Preprint. Available at `http://www.eurandom.tue.nl/Past\%20years/reports/2002/030jsreport.pdf`

Beirlant, J., Joossens, E. and Segers, J. (2004a). Discussion of "Generalized Pareto Fit to the Society of Actuaries' Large Claims Database" by A. Cebrián, M. Denuit and Ph. Lambert. *N. Am. Actuar. J.* **8** 108-111.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D. and Ferro, C. (2004b). *Statistics of Extremes: Theory and Applications.* Wiley, New York.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

Cohen, J.P. (1982). The penultimate form of approximation to normal extremes. *Adv. Appl. Probab.* **14** 324-339.

Embrechts, P.A.L., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27** 760-781.

de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc.* **61** 381-395.