QQ PLOTS, RANDOM SETS
AND DATA FROM A HEAVY TAILED DISTRIBUTION

BIKRAMJIT DAS AND SIDNEY I. RESNICK

ABSTRACT. The QQ plot is a commonly used technique for informally deciding whether a univariate random sample of size n comes from a specified distribution F. The QQ plot graphs the sample quantiles against the theoretical quantiles of F and then a visual check is made to see whether or not the points are close to a straight line. For a location and scale family of distributions, the intercept and slope of the straight line provide estimates for the shift and scale parameters of the distribution respectively. Here we consider the set $\mathcal{S}_n$ of points forming the QQ plot as a random closed set in $\mathbb{R}^2$. We show that under certain regularity conditions on the distribution F, $\mathcal{S}_n$ converges in probability to a closed, non-random set. In the heavy tailed case where $1 - F$ is a regularly varying function, a similar result can be shown but a modification is necessary to provide a statistically sensible result since typically F is not completely known.

1. INTRODUCTION

Given a random sample of univariate data points, a pertinent question is whether this sample comes from some specified distribution F. A variant question is whether the sample is from a location/scale family derived from F. Decision techniques are based on how close the empirical distribution of the sample and the distribution F are for some sample size n. The empirical distribution function of the iid random variables $X_1, \ldots, X_n$ is

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x), \qquad -\infty < x < \infty.$$

The Kolmogorov-Smirnov (KS) statistic is one way to measure the distance between the empirical distribution function and the distribution function F. Glivenko and Cantelli showed (see, for example, Serfling (1980)) that the KS-statistic converges to 0 almost surely. The QQ (or quantile-quantile) plot is another commonly used device to graphically, quickly and informally test the goodness-of-fit of a sample in an exploratory way. It has the advantage of being a graphical tool, which is visually appealing and easy to understand. The QQ plot measures how close the sample quantiles are to the theoretical quantiles. For $0 < p < 1$, the pth *quantile* of F is defined by

$$(1.1) \qquad F^{\leftarrow}(p) := \inf\{x : F(x) \geq p\}.$$

The sample p^{th} quantile can be similarly defined as $F_n^{\leftarrow}(p)$. If $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics from the sample, then $F_n^{\leftarrow}(p) = X_{\lceil np \rceil:n}$, where as usual $\lceil np \rceil$ is the smallest integer greater than or equal to np. For $0 < p < 1$, $X_{\lceil np \rceil:n}$ is a strongly consistent estimator of $F^{\leftarrow}(p)$ (Serfling, 1980, page 75).

Rather than considering individual quantiles, the QQ plot considers the sample as a whole and plots the sample quantiles against the theoretical quantiles of the specified target distribution F. If we have a correct target distribution, the QQ plot hugs a straight line through the the origin at an angle of $45°$. Sometimes we have a location and scale family correctly specified up to unspecified location and scale and in such cases, the QQ plot concentrates around a straight line with some slope (not necessarily $45°$) and intercept (not necessarily 0); the slope and intercept estimate the scale and location. Since a variety of estimation and inferential procedures in the practice of statistics depends on the assumption of normality of the data, the normal QQ plot is one of the most commonly used.

Our goal here is to formally prove the convergence of the QQ plot (perhaps suitably modified) to a straight line. This would show the asymptotic consistency of the QQ plot. The QQ plot formed by a sample of size n can be considered a closed subset of $\mathbb{R}^2$ denoted by $\mathcal{S}_n$. This set of points that form the QQ plot in $\mathbb{R}^2$ is

$$(1.2) \qquad \mathcal{S}_n := \{(F^{\leftarrow}(\frac{i}{n+1}), X_{i:n}), \ 1 \leq i \leq n\}$$

where the function $F^{\leftarrow}(\cdot)$ is defined by (1.1). For each n, $\mathcal{S}_n$ is a random closed set. Note that, if $\{\mathcal{S}_n\}$ has an almost sure limit S then this limit set by the Hewitt-Savage $0-1$ law must be almost surely constant. A straight line (or some closed subset of a straight line) is also a closed set in $\mathbb{R}^2$. Under certain regularity conditions on F, we show that the random set $\mathcal{S}_n$ converges in probability to a straight line (or some closed subset of a straight line), in a suitable topology on closed subsets of $\mathbb{R}^2$.

Section 2 is devoted to preliminary results on the convergence of random closed sets. We also discuss a result on convergence of quantiles and, because of our interest in heavy tails, we introduce the concept of regular variation. In Section 3, we assume the random variables have a specified distribution F and we consider convergence of the random closed sets $\mathcal{S}_n$ forming the QQ plot. In Section 4, the idea of the QQ plot is extended to the case where we know that the data is heavy tailed, that is $1 - F$ is regularly varying. We assume we do not know the exact distribution of F; we presume the distribution is heavy tailed but do not know either the tail index or the slowly varying component. The usual QQ plot is not informative in a statistical sense and hence must be modified by a thresholding technique.

In Corollary 3.4 we have convergence of a log-transformed version of the QQ plot to a straight line when the distribution of the random sample is Pareto. Now Pareto being a special case of a distribution with regularly varying tail, we use the same plotting technique for random variables having a regularly varying tail after thresholding the data. We provide a convergence in probability result considering the $k = k(n)$ upper order statistics of the data set where $k \to \infty$ and $k/n \to 0$. In Section 5, a continuity result is provided for a least squares line through these special kinds of closed sets. See Beirlant et al. (1996), Kratz and Resnick (1996).

2. Preliminaries

2.1. **Closed sets and the Fell topology.** We denote the distance between the points $\boldsymbol{x}$ and $\boldsymbol{y}$ by $d(\boldsymbol{x}, \boldsymbol{y})$; $\mathcal{F}, \mathcal{G}$ and $\mathcal{K}$ are the classes of closed, open and compact subsets of $\mathbb{R}^d$ respectively. These quantities are sometimes subscripted by the dimension of the space if this needs to be emphasized for clarity. We are interested in closed sets because the sets of interest such as $\mathcal{S}_n$ are random closed sets. There are several ways to define a topology on the space of closed sets. The Vietoris topology and the Fell topology are frequently used and these are hit-or-miss kinds of topologies. We shall discuss the Fell topology below. For further discussion refer to Beer (1993), Matheron (1975), Molchanov (2005).

For a set $B \subset \mathbb{R}^d$, define $\mathcal{F}_B$ as the class of closed sets hitting B and $\mathcal{F}^B$ as the class of closed sets disjoint from B:

$$\mathcal{F}_B = \{F : F \in \mathcal{F}, \ F \bigcap B \neq \emptyset\}, \quad \mathcal{F}^B = \{F : F \in \mathcal{F}, \ F \bigcap B = \emptyset\}.$$

Now the space $\mathcal{F}$ can be topologized by the Fell topology which has as its subbase the families $\{\mathcal{F}^K, K \in \mathcal{K}\}$ and $\{\mathcal{F}_G, G \in \mathcal{G}\}$.

A sequence $\{F_n\}$ converges in the Fell topology towards a limit F in $\mathcal{F}$ (written $F_n \to F$) if and only if it satisfies two conditions:

    (1) If an open set G hits F, G hits all F_n, provided n is sufficiently large.
    (2) If a compact set K is disjoint from F, it is disjoint from F_n for all sufficiently large n.

The following result (Matheron, 1975) provides useful conditions for convergence.

Lemma 2.1. *For $F_n, F \in \mathcal{F}, n \geq 1$, $F_n \to F$ as $n \to \infty$ if and only if the following two conditions hold*

$$(2.1) \qquad \text{For any } \boldsymbol{y} \in F, \text{ for all large } n, \text{ there exists } \boldsymbol{y}_n \in F_n \text{ such that } d(\boldsymbol{y}_n, \boldsymbol{y}) \to 0 \text{ as } n \to \infty.$$

$$(2.2) \qquad \text{For any subsequence } \{n_k\}, \text{ if } \boldsymbol{y}_{n_k} \in F_{n_k} \text{ converges, then } \lim_{k \to \infty} \boldsymbol{y}_{n_k} \in F.$$

Furthermore, convergence of sets $\mathcal{S}_n \to \mathcal{S}$ in $\mathcal{K}$ is equivalent to the analogues of (2.1) and (2.2) holding as well as $\sup_{j \geq 1} \sup\{\|\boldsymbol{x}\| : \boldsymbol{x} \in S_j\} < \infty$ for some norm $\| \cdot \|$ on $\mathbb{R}^d$.

Note that if the sets are random elements of $\mathcal{K}$ and $S \in \mathcal{K}$ is non-random, then Lemma 2.1 can be used to characterize almost sure convergence or convergence in probability. We are going to define random sets in the next subsection.

Definition 2.1 (Hausdorff Metric)**.** Suppose $d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ is a metric on $\mathbb{R}^d$. Then for $S, T \in \mathcal{K}$, define the Hausdorff metric (Matheron, 1975) $D : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ by

$$(2.3) \qquad\qquad D(S, T) = \inf\{\delta : S \subset T^\delta, T \subset S^\delta\},$$

where for $S \in \mathcal{K}$ and $\delta > 0$,

$$(2.4) \qquad\qquad S^\delta = \{\boldsymbol{x} : d(\boldsymbol{x}, \boldsymbol{y}) < \delta \text{ for some } \boldsymbol{y} \in S\}$$

is the δ-neighborhood or δ-swelling of S.

The topology usually used on $\mathcal{K}$ is the *myopic topology* with sub-base elements $\{\mathcal{K}^F, F \in \mathcal{F}\}$ and $\{\mathcal{K}_G, G \in \mathcal{G}\}$. The myopic topology on $\mathcal{K}$ is stronger than the Fell topology relativized to $\mathcal{K}$. The topology on $\mathcal{K}' = \mathcal{K} \setminus \{\emptyset\}$ generated by the Hausdorff metric is equivalent to the myopic topology on $\mathcal{K}'$ (Molchanov, 2005, page 405).

In certain cases, convergence on $\mathcal{F}$ can be reduced to convergence on $\mathcal{K}$.

Lemma 2.2. *Suppose F_n and F are closed sets in $\mathcal{F}$ and and that there exist $\mathcal{K}_1 \subset \mathcal{K}$ satisfying*

(1) $\displaystyle\bigcup_{K \in \mathcal{K}_1} K = \mathbb{E}.$

(2) *For $\delta > 0$ and $K \in \mathcal{K}$, we have $\overline{K^\delta} \in \mathcal{K}_1$.*

(3) $F_n \bigcap K \to F \bigcap K, \quad \forall K \in \mathcal{K}_1.$

Then $F_n \to F$ in $\mathcal{F}$.

Remark 2.1. The converse is false. Let $\mathbb{E} = \mathbb{R}$, $F_n = \{1/n\}$, $F = \{0\}$ and $K = [-1, 0]$. Then $F_n \to F$ but

$$F_n \bigcap K = \emptyset \nrightarrow F \bigcap K = F.$$

The operation of intersection is not a continuous operation in $\mathcal{F} \times \mathcal{F}$ (Molchanov, 2005, page 400); it is only upper semicontinuous (Matheron, 1975, page 9).

Proof. We use Lemma 2.1. If $x \in F$, there exists $K \in \mathcal{K}_1$ and $x \in K$. So $x \in F \cap K$ and from Lemma 2.1, since $F_n \cap K \to F \cap K$ as $n \to \infty$, we have existence of $x_n \in F_n \cap K$ and $x_n \to x$. So we have produced $x_n \in F_n$ and $x_n \to x$ as required for (2.1).

To verify (2.2), suppose $\{x_{n_k}\}$ is a subsequence such that $x_{n_k} \in F_{n_k}$ and $\{x_{n_k}\}$ converges to, say, x_∞. We need to show $x_\infty \in F$. There exists $K_\infty \in \mathcal{K}_1$ such that $x_\infty \in K_\infty$. For any $\delta > 0$, $x_{n_k} \in \overline{K_\infty^\delta} \in \mathcal{K}_1$ for all sufficiently large n_k. So $x_{n_k} \in F_{n_k} \cap \overline{K_\infty^\delta}$. Since $F_{n_k} \cap \overline{K_\infty^\delta} \to F \cap \overline{K_\infty^\delta}$, we have $\lim_{k \to \infty} x_{n_k} = x_\infty \in F \cap \overline{K_\infty^\delta}$. So $x \in F$. $\qquad\square$

The next result shows when a point set approximating a curve actually converges to the curve. For this Lemma, $C(0, 1]$ is the class of real valued continuous functions on $(0, 1]$ and $D_l(0, \infty]$ is the class of left continuous functions on $(0, \infty]$ with finite right hand limits.

Lemma 2.3. *Suppose $0 \leq x(\cdot) \in C(0, 1]$ is continuous on $(0, 1]$ and strictly decreasing with $\lim_{\epsilon \downarrow 0} x(\epsilon) = \infty$. Suppose further that $y_n(\cdot) \in D_l(0, 1]$ and $y(\cdot) \in C(0, 1]$ and $y_n \to y$ locally uniformly on $(0, 1]$; that is, uniformly on compact subintervals bounded away from 0. Then for $k = k(n) \to \infty$,*

$$F_n := \left\{\left(x(\tfrac{j}{k}), y_n(\tfrac{j}{k})\right); 1 \leq j \leq k\right\} \to F := \left\{\left(x(t), y(t)\right); 0 < t \leq 1\right\} = \left\{\left(u, y(x^\leftarrow(u))\right); x(1) \leq u < \infty\right\},$$

in $\mathcal{F}$.

Proof. Pick $t \in (0, 1]$, so that $(x(t), y(t)) \in F$. Then

$$F_n \ni \big(x(\lceil kt \rceil/k), y_n(\lceil kt \rceil/k)\big) \to \big(x(t), y(t)\big) \in F,$$

in $\mathbb{R}^2$, verifying (2.1). For (2.2), Suppose $\big(x(j(n')/k(n'), y_{n'}(j(n')/k(n'))\big) \in F_{n'}$ is a convergent subsequence in $\mathbb{R}^2$. Then $\{x(j(n')/k(n'))\}$ is convergent in $\mathbb{R}$ and because $x(\cdot)$ is strictly monotone, $\{j(n')/k(n')\}$ converges to some $l \in (0, 1]$. Then

$$F_{n'} \ni \big(x(j(n')/k(n'), y_{n'}(j(n')/k(n'))\big) \to \big(x(l), y(l)\big) \in F,$$

which verifies (2.2). $\square$

2.2. Random closed sets and weak convergence.
In this section, we review definitions and characterizations of weak convergence of random closed sets. In subsequent sections we will show convergence in probability, but since the limit sets will be non-random, weak convergence and convergence in probability coincide. See also Matheron (1975), Molchanov (2005).

Let $(\Omega, \mathcal{A}, P')$ be a complete probability space. $\mathcal{F}$ is the space of all closed sets in $\mathbb{R}^d$ topologized by the Fell topology. Let $\sigma_{\mathcal{F}}$ denote the Borel σ-algebra generated by the Fell topology of open sets. A *random closed set* $X : \Omega \mapsto \mathcal{F}$ is a measurable mapping from $(\Omega, \mathcal{A}, P')$ to $(\mathcal{F}, \sigma_{\mathcal{F}})$. Denote by P the induced probability on $\sigma_{\mathcal{F}}$, that is, $P = P' \circ X^{-1}$. A sequence of random closed sets $\{X_n\}_{n \geq 1}$ weakly converges to a random closed set X with distribution P) if the corresponding induced probability measures $\{P_n\}_{n \geq 1}$ converge weakly to P, i.e.,

$$P_n(\mathcal{B}) = P'_n \circ X_n^{-1}(\mathcal{B}) \to P(\mathcal{B}) = P' \circ X^{-1}(\mathcal{B}), \qquad \text{as } n \to \infty,$$

for each $\mathcal{B} \in \sigma_{\mathcal{F}}$ such that $P(\partial \mathcal{B}) = 0$.

This is not always straightforward to verify from the definition. We find useful the following characterization of weak convergence in terms of sup-measures (Vervaat, 1997). Suppose $h : \mathbb{R}^d \mapsto \mathbb{R}_+ = [0, \infty)$. For $X \subset \mathbb{R}^d$, define $h(X) = \{h(x) : x \in X\}$ and $h^\vee$ is the sup-measure generated by h defined by

$$h^\vee(X) = \sup\{h(x) : x \in X\}$$

(Molchanov, 2005, Vervaat, 1997). These definitions permit the following characterization (Molchanov, 2005, page 87).

Lemma 2.4. *A sequence $\{X_n\}_{n \geq 1}$ of random closed sets converges weakly to a random closed set X if and only if $\mathbb{E}h^\vee(X_n)$ converges to $\mathbb{E}h^\vee(X)$ for every non-negative continuous function $h : \mathbb{R}^d \mapsto \mathbb{R}$ with a bounded support.*

2.3. Convergence of sample quantiles.
The sample quantile is a strongly consistent estimator of the population quantile (Serfling (1980), page 75). The weak consistency of sample quantiles as estimators of population quantiles was shown by Smirnov (1949); see also (Resnick, 1999, page 179). We will make use of the Glivenko-Cantelli lemma describing uniform convergence of the sample empirical distribution and also take note of the following quantile estimation result.

Lemma 2.5. *Suppose F is strictly increasing at $F^\leftarrow(p)$ which means that for all $\epsilon > 0$,*

$$F(F^\leftarrow(p - \epsilon)) < p < F(F^\leftarrow(p + \epsilon)).$$

Then we have that the pth sample quantile, $X_{\lceil np \rceil : n}$ is a weakly consistent quantile estimator,

$$X_{\lceil np \rceil : n} \xrightarrow{P} F^\leftarrow(p)$$

As before, $\lceil np \rceil$ is the 1st integer $\geq np$ and $X_{i:n}$ is the ith smallest order statistic.

2.4. **Regular variation.** Regular variation is the mathematical underpinning of heavy tail analysis. It is discussed in many books such as Bingham et al. (1987), de Haan (1970), de Haan and Ferreira (2006), Geluk and de Haan (1987), Resnick (1987, 2006), Seneta (1976).

Definition 2.2 (Regular variation)**.** A measurable function $U(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ if for $x > 0$

$$(2.5) \qquad\qquad \lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^\rho.$$

We write $U \in RV_\rho$.

Remark 2.2. When $\rho = 0$ we call $U(\cdot)$ *slowly varying* and denote it by $L(\cdot)$. For $\rho \in \mathbb{R}$, we can always write $U \in RV_\rho$ as:

$$(2.6) \qquad\qquad U(x) = x^\rho L(x)$$

where $L(\cdot)$ is slowly varying.

3. QQ plots from a known distribution: Random sets converging to a constant set

In this section, we will use the results in Section 2 to show the convergence of the random closed sets given by (1.2) consisting of the points forming the QQ plot to a non-random set in $\mathbb{R}^2$. First we consider the easiest case where the random variables are iid from a uniform distribution. Then we consider more general distributions which are continuous and strictly increasing on their support. This result will be derived from the uniform case. Because we are interested in heavy tailed distributions, our final corollary in this section is about the Pareto distribution which is the exemplar of the heavy tailed distribution.

3.1. **The Uniform case.** The first simple example is QQ plot from the uniform distribution.

Proposition 3.1. *Suppose $U_1, U_2, \ldots, U_n$ are iid U(0,1). Denote the order statistics of this sample by $U_{1:n} \le U_{2:n} \le \ldots \le U_{n:n}$. Define*

$$(3.1) \qquad\qquad \mathcal{S}_n := \{(\frac{i}{n+1}, U_{i:n}),\ 1 \le i \le n\}$$

and

$$(3.2) \qquad\qquad \mathcal{S} := \{(x, x) : 0 \le x \le 1\}.$$

Then $\mathcal{S}_n \overset{a.s.}{\to} \mathcal{S}$ in $\mathcal{K}_2$.

Proof. We apply the convergence criterion given in Lemma 2.1. The empirical distribution $U_n(x) = n^{-1} \sum_{i=1}^n I(U_i \le x)$ converges uniformly for almost all sample paths to x, $0 \le x \le 1$. Without loss of generality suppose this true for all sample paths. Then for all sample paths, the same is true for the inverse process $U_n^\leftarrow(p) = U_{\lceil np \rceil : n}$, $0 \le p \le 1$; that is

$$\sup_{0 \le p \le 1} |U_{\lceil np \rceil : n} - p| \to 0, \quad (n \to \infty).$$

Pick $0 \le y \le 1$ and let $\boldsymbol{y} = (y, y) \in \mathcal{S}$. For each n, define $\boldsymbol{y}_n$ by

$$(3.3) \qquad\qquad \boldsymbol{y}_n = \Big(\frac{\lceil ny \rceil}{n+1}, U_{\lceil ny \rceil : n} \Big),$$

so that $\boldsymbol{y}_n \in \mathcal{S}_n$. Since $|ny - \lceil ny \rceil| \le 1$, $\lceil ny \rceil / n + 1 \to y$ and since $U_{\lceil ny \rceil : n} \to y$, we have $\boldsymbol{y}_n \to (y, y) \in \mathcal{S}$. Hence criterion (2.1) from Lemma 2.1 is satisfied.

Now suppose we have a subsequence $\{n_k\}$ such that $\boldsymbol{y}_{n_k} \in \mathcal{S}_{n_k}$ converges. Then $\boldsymbol{y}_{n_k}$ is of the form $\boldsymbol{y}_{n_k} = (i_{n_k}/(n_k + 1), U_{i_{n_k} : n_k})$ for some $1 \le i_{n_k} \le n$ and for some $x \in [0,1]$, we have $i_{n_k}/(n_k + 1) \to x$ and hence also $i_{n_k}/n_k \to x$. This implies

$$U_{i_{n_k} : n_k} = U_{\lceil n_k \cdot \frac{i_{n_k}}{n_k} \rceil : n_k} \to x,$$

and therefore $\boldsymbol{y}_{n_k} \to (x, x)$ as required for (2.2). □

3.2. **Convergence for more general distributions.** Now consider a distribution function F which is more general than the uniform, assuming that F is strictly increasing and continuous on its support so that $F^{\leftarrow}$ is unique.

Proposition 3.2. *Suppose $X_1, \ldots, X_n$ are iid with common distribution $F(\cdot)$ and $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics from this sample. If F is strictly increasing and continuous on its support, then*

$$\mathcal{T}_n := \{(F^{\leftarrow}(\frac{i}{n+1}), X_{i:n}); 1 \leq i \leq n\}$$

converges in probability to

$$\mathcal{T} := \{(x, x); x \in support(F)\}$$

in $\mathcal{F}_2$.

Proof. According to Lemma 2.4, we must prove for any non-negative continuous $h : \mathbb{R}^2 \mapsto \mathbb{R}_+$ with compact support that as $n \to \infty$,

$$\mathbb{E}(h^{\vee}(\mathcal{T}_n)) \to \mathbb{E}(h^{\vee}(\mathcal{T})).$$

Since F is continuous, $F(X_1), F(X_2), \ldots, F(X_n)$ are iid and uniformly distributed on $[0, 1]$. Therefore from Proposition 3.1 we have that

$$(3.4) \qquad \mathcal{S}_n := \{(\frac{i}{n+1}, F(X_{i:n})); 1 \leq i \leq n\} \stackrel{d}{=} \{\frac{i}{n+1}, U_{i:n}); 1 \leq i \leq n\} \stackrel{a.s.}{\to} \mathcal{S} = \{(x, x); 0 \leq x \leq 1\}$$

in $\mathcal{K}_2$.

We now proceed by considering cases which depend on the nature of the support of F. We will need the following identity. For any closed set X, function $f : \mathbb{R}^2 \mapsto \mathbb{R}_+$ and function $\psi : \mathbb{R}^2 \mapsto \mathbb{R}^2$, we have,

$$(3.5) \qquad f^{\vee} \circ \psi(X) = \sup_{t \in \psi(X)} f(t) = \sup_{s \in X} f(\psi(s)) = \sup_{s \in X} f \circ \psi(s) = (f \circ \psi)^{\vee}(X).$$

Case 1: The support of F is compact, say $[a, b]$. This implies $F^{\leftarrow}(0) = a$, $F^{\leftarrow}(1) = b$. Define the map $g : [0, 1]^2 \mapsto [a, b]^2$ by

$$g(x, y) = (F^{\leftarrow}(x), F^{\leftarrow}(y)).$$

Since F is strictly increasing, observe that $g(\mathcal{S}_n) = \mathcal{T}_n$ and $g(\mathcal{S}) = \mathcal{T}$. Define $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ as the extension of g to all of $\mathbb{R}^2$:

$$g^*(x, y) = (g_1(x), g_1(y))$$

where

$$g_1(z) = \begin{cases} F^{\leftarrow}(z), & 0 \leq z \leq 1 \\ a, & z \leq 0 \\ b, & z \geq 1. \end{cases}$$

This makes $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ continuous. Since both $\mathcal{S}_n$ and $\mathcal{S}$ are subsets of $[0, 1] \times [0, 1]$, we have $g(\mathcal{S}_n) = g^*(\mathcal{S}_n)$ and $g(\mathcal{S}) = g^*(\mathcal{S})$. Let f be a continuous function on $\mathbb{R}^2$ with bounded support and we have, as $n \to \infty$, using (3.5),

$$\mathbb{E}f^{\vee}(\mathcal{T}_n) = \mathbb{E}f^{\vee}(g(\mathcal{S}_n)) = \mathbb{E}f^{\vee}(g^*(\mathcal{S}_n))$$
$$= \mathbb{E}(f \circ g^*)^{\vee}(\mathcal{S}_n) \to \mathbb{E}(f \circ g^*)^{\vee}(\mathcal{S}).$$

The previous convergence results from $f \circ g^* : \mathbb{R}^2 \mapsto \mathbb{R}_+$ being continuous with bounded support, $\mathcal{S}_n \stackrel{P}{\to} \mathcal{S}$, and Lemma 2.4. The term to the right of the convergence arrow above equals

$$= \mathbb{E}f^{\vee}(g^*(\mathcal{S})) = \mathbb{E}f^{\vee}(g(\mathcal{S})) = \mathbb{E}f^{\vee}(\mathcal{T}).$$

Therefore $\mathcal{T}_n$ converges to $\mathcal{T}$ weakly and since $\mathcal{T}$ is a non-random set, this convergence is also true in probability.

Case 2: The support of F is $\mathbb{R} = (-\infty, \infty)$. Now define $g : (0,1)^2 \mapsto \mathbb{R}^2$ by

$$g(x,y) = (F^{\leftarrow}(x), F^{\leftarrow}(y)).$$

Since F is strictly increasing, $g(\mathcal{S}_n) = \mathcal{T}_n$ and $g(\mathcal{S} \cap (0,1)^2) = \mathcal{T}$. Let f be a continuous function with bounded support in $[-M, M]^2$, for some $M > 0$. Extend the definition of g to all of $\mathbb{R}^2$ by defining $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ as

$$g^*(x,y) = (g_1(x), g_1(y)),$$

where

$$g_1(z) = \begin{cases} F^{\leftarrow}(z), & -M \leq F^{\leftarrow}(z) \leq M, \\ F^{\leftarrow}(-M), & F^{\leftarrow}(z) \leq -M, \\ F^{\leftarrow}(M), & F^{\leftarrow}(z) \geq M. \end{cases}$$

Therefore $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ is continuous. Now note that since $support(f) \subseteq [-M, M]^2$ and $g(x,y) = g^*(x,y)$ for $(x,y) \in [-M, M]^2$, we will have $f \circ g = f \circ g^*$. Therefore

$$\mathbb{E}f^{\vee}(\mathcal{T}_n) = \mathbb{E}f^{\vee}(g(\mathcal{S}_n)) = \mathbb{E}(f \circ g)^{\vee}(\mathcal{S}_n) = \mathbb{E}(f \circ g^*)^{\vee}(\mathcal{S}_n) \to \mathbb{E}(f \circ g^*)^{\vee}(\mathcal{S}).$$

As with Case 1, the convergence follows from $f \circ g^* : \mathbb{R}^2 \mapsto \mathbb{R}_+$ being continuous with bounded support, $\mathcal{S}_n \xrightarrow{P} \mathcal{S}$ and Choquet's theorem 2.4. The term to the right of the convergence arrow equals

$$= \mathbb{E}(f \circ g)^{\vee}(\mathcal{S}) = \mathbb{E}f^{\vee}(g(\mathcal{S})) = \mathbb{E}f^{\vee}(\mathcal{T}).$$

Therefore $\mathcal{T}_n$ converges to $\mathcal{T}$ weakly. But since $\mathcal{T}$ is a non-random set, this convergence is true also in probability.

Case 3: The support of F is of the form $[a, \infty)$ or $(-\infty, b]$. This case can be examined in a similar manner as we have done for Cases 1 and 2 by considering each end-point of the interval of support of F according to its nature. $\square$

Corollary 3.3. *If F is exponential with parameter $\alpha > 0$, i.e., $F(x) = 1 - e^{-\alpha x}$, $x > 0$, we have*

$$\{(-\frac{1}{\alpha} \log(1 - \frac{i}{n+1}), X_{i:n}); 1 \leq i \leq n\} \xrightarrow{P} \{(x,x) : 0 \leq x < \infty\}.$$

Corollary 3.4. *If F is Pareto with parameter $\alpha > 0$, i.e., $F(x) = 1 - x^{-\alpha}$, $x > 1$, we have*

$$\{(-\log(1 - \frac{i}{n+1}), \log X_{i:n}); 1 \leq i \leq n\} \xrightarrow{P} \{(x, \frac{x}{\alpha}) : 0 \leq x < \infty\}.$$

4. QQ plots: Convergence of random sets in the regularly varying case

The classical QQ plot can be graphed only if we know the target distribution F at least up to location and scale. We would like to extend the idea of QQ plots to the case where the data is from a heavy tailed distribution; this is a semi-parametric assumption which is more general than assuming the target distribution F is known up to location and scale.

We model a one-dimensional heavy-tailed distribution function F by assuming it has a regularly varying tail with some index $-\alpha$, for $\alpha > 0$; that is, if X has distribution F then,

$$(4.1) \qquad P[X > x] = 1 - F(x) = \bar{F}(x) = x^{-\alpha} L(x), \qquad x > 0$$

where L is slowly varying. In at least an exploratory context, how can the QQ plot be used to validate this assumption and also to estimate α? (See Resnick (2006, page 106).)

Notice that if we take $L \equiv 1$, F turns out to be a Pareto distribution with parameter α. In Corollary (3.4), we have seen that if F has a Pareto distribution with parameter α, then $\mathcal{S}_n$ defined as:

$$(4.2) \qquad \mathcal{S}_n := \{(-\log(1 - \frac{i}{n+1}), \log X_{i:n}); 1 \leq i \leq n\}$$

converges in probability to the set

$$(4.3) \qquad \mathcal{S} = \{(x, \frac{x}{\alpha}); 0 \le x < \infty\}.$$

Keeping this in mind, when we have a general $\bar{F} \in RV_{-\alpha}$, let us define $\mathcal{S}_n$ exactly as in (4.2). Then we are able to show that, $\mathcal{S}_n$ converges in probability to the set

$$(4.4) \qquad \mathcal{S} = \{(\alpha x, x + \frac{1}{\alpha} \log L(F^{\leftarrow}(1 - e^{-\alpha x}))); \; 0 \le x < \infty\}.$$

But, since we do not know the slowly varying function $L(\cdot)$, this result is not useful for inference purposes. Estimating α from such a set is not possible unless $L(\cdot)$ is known, nor is it clear how $\mathcal{S}_n$ graphically approximating such a set would allow us to validate the model assumption of a regularly varying tail.

Consequently we concentrate on a different asymptotic regime where the asymptotic behavior of the random closed set can be freed from $L(\cdot)$. For a sample of size n from the distribution F, where $\bar{F} \in RV_{-\alpha}$, we consider the upper $k = k(n)$ order statistics of the sample where $k(n)/n \to 0$ and construct a QQ plot similar to (4.2). We assume that $d_{\mathcal{F}}(\cdot, \cdot)$ is some metric on $\mathcal{F}$ which is compatible with the Fell topology. Note Flachsmeyer (1963/1964) characterized the metrizability of the Fell topology and since $\mathbb{R}^d$ is locally compact, Hausdorff and second countable his results apply and allow the conclusion that $\mathcal{F}$ is metrizable under the Fell topology.

For what follows, when $A \in \mathcal{F}_2$, we write $A + (t_1, t_2) = \{\boldsymbol{a} + (t_1, t_2) : \boldsymbol{a} \in A\}$ for the translation of A.

Proposition 4.1. *Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from F where $\bar{F} \in RV_{-\alpha}$ and $X_{(1)} \ge X_{(2)} \ge \ldots \ge X_{(n)}$ are the order statistics in decreasing order. Define*

$$\mathcal{S}_n = \{(-\log \frac{j}{n+1}, \log X_{(j)}); j = 1, \ldots, k\}$$

where $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Also define

$$\mathcal{T}_n = \{(x, \frac{x}{\alpha}); x \ge 0\} + (-\log \frac{k}{n+1}, \log X_{(k)})$$

Then as $n \to \infty$

$$d_{\mathcal{F}}(\mathcal{S}_n, \mathcal{T}_n) \xrightarrow{P} 0$$

Remark 4.1. So after a logarithmic transformation of the data, we make the QQ plot by only comparing the k largest order statistics with the corresponding theoretical exponential distribution quantiles. This produces an asymptotically linear plot of slope $1/\alpha$ starting from the point $(-\log \frac{k}{n+1}, \log X_{(k)})$.

Proof. Define

$$\mathcal{S}'_n = \{(-\log \frac{j}{k}, \log \frac{X_{(j)}}{X_{(k)}}); 1 \le j \le k\}, \text{ and } \mathcal{T} = \{(x, \frac{x}{\alpha}); 0 \le x < \infty\}.$$

Note that we can write

$$\mathcal{S}'_n = \{(-\log \frac{j}{k}, \log \frac{X_{(j)}}{X_{(k)}}); 1 \le j \le k\} = \{(-\log t, \log \frac{X_{([kt])}}{X_{(k)}}); t \in \{\frac{1}{k}, \ldots, \frac{k-1}{k}, 1\}\},$$

and also write $\mathcal{T}$ as

$$\mathcal{T} = \{(x, \frac{x}{\alpha}); x \ge 0\} = \{(-\log t, -\frac{1}{\alpha} \log t); 0 < t \le 1\},$$

where we put $x = -\log t$. We first show $\mathcal{S}'_n \xrightarrow{P} \mathcal{T}$.

Referring to Lemma 2.3, set

$$x(t) = -\log t, \quad Y_n(t) = \log \frac{X_{(\lceil kt \rceil)}}{X_{(k)}}, \quad y(t) = -\frac{1}{\alpha} \log t, \qquad 0 < t \le 1.$$

¿From Resnick (2006, page 82, equation (4.18)), we have $Y_n \xrightarrow{P} y$, in $D_l(0, 1]$, the left continuous functions on $(0, 1]$ with finite right limits, metrized by the Skorohod metric. Suppose $\{n''\}$ is a subsequence. There

exists a further subsequence $\{n'\} \subset \{n''\}$ such that $Y_{n'} \overset{a.s.}{\to} y$, in $D_l(0,1]$, and by Lemma 2.3, $\mathcal{S}'_{n'} \overset{a.s.}{\to} \mathcal{T}$ in $\mathcal{F}$. Therefore $\mathcal{S}'_n \overset{P}{\to} \mathcal{T}$, in $\mathcal{F}$, as $n \to \infty$.

Now observe that with $\boldsymbol{a}_n := (-\log \frac{k}{n+1}, \log X_{(k)})$, we have

$$
\begin{aligned}
\mathcal{S}_n &= \{(-\log \frac{j}{n+1}, \log X_{(j)}); j = 1, \ldots, k\} \\
&= \{(-\log \frac{j}{k}, \log \frac{X_{(j)}}{X_{(k)}}); j = 1, \ldots, k\} + (-\log \frac{k}{n+1}, \log X_{(k)}) \\
&= \mathcal{S}'_n + \boldsymbol{a}_n.
\end{aligned}
$$

Also,

$$
\mathcal{T}_n = \{(x, \frac{x}{\alpha}); x \geq 0\} + (-\log \frac{k}{n+1}, \log X_{(k)}) = \mathcal{T} + \boldsymbol{a}_n.
$$

Now, since $d_{\mathcal{F}}(\mathcal{S}'_n, \mathcal{T}) \overset{P}{\to} 0$, we get

$$
d_{\mathcal{F}}(\mathcal{S}_n, \mathcal{T}_n) = d_{\mathcal{F}}(\mathcal{S}'_n + \boldsymbol{a}_n, \mathcal{T} + \boldsymbol{a}_n) = d_{\mathcal{F}}(\mathcal{S}'_n, \mathcal{T}) \overset{P}{\to} 0,
$$

as required. □

5. LEAST SQUARES LINE THROUGH A CLOSED SET

5.1. **Convergence of the least squares line.** The previous two sections gave results about the convergence of the QQ plot to a straight line in the Fell topology of $\mathcal{F}_2$. It is of interest to know whether some functional of closed sets is continuous or not and, in particular, the slope of the least squares line through the points of QQ plot is one such functional. The slope of the least squares line is an estimator of scale for location/scale families and this leads to an estimate of the heavy tail index α; see Beirlant et al. (1996), Kratz and Resnick (1996) and Resnick (2006, Section 4.6).

Intuition suggests that when a sequence of finite sets converges to a line, the slope of the least squares line should converge to the slope of the limiting line. However there are subtleties which prevent this from being true in general. We need some restriction on the point sets that converge, since otherwise, a sequence of point sets which are essentially linear except for a vanishing bump, may converge to a line but the bump may skew the least squares line sufficiently to prevent the slope from converging; see Example 5.1 below.

The following Proposition provides a condition for the continuity property to hold. First define the subclass $\mathcal{F}_{\text{finite or line}} \subset \mathcal{F}_2$ to be the closed sets of $\mathcal{F}_2$ which are either sets of finite cardinality or closed, bounded line segments. These are the only cases of compact sets where it is clear how to define a least squares line. For $F \in \mathcal{F}_{\text{finite or line}}$, the functional LS is defined in the obvious way:

$$LS(F) = \text{slope of the least squares line through the closed set } F$$

For the next proposition, we consider sets $F_n := \{(x_i(n), y_i(n)) : 1 \leq i \leq k_n\}$ of points and write $\bar{x}_n = \sum_{j=1}^{k_n} x_j(n)/k_n$ and $\bar{y}_n = \sum_{j=1}^{k_n} y_j(n)/k_n$. Also, for a finite set $\mathcal{S}_n$, $\#\mathcal{S}_n$ denotes the cardinality of $\mathcal{S}_n$.

Proposition 5.1. *Suppose we have a sequence of sets $F_n := \{(x_i(n), y_i(n)) : 1 \leq i \leq k_n\} \in \mathcal{K}_2$, each consisting of k_n points, which converge to a bounded line segment $F \in \mathcal{K}_2$ with slope m where $|m| < \infty$, as $k_n \to \infty$. Then*

$$LS(F_n) \to LS(F) = m$$

provided the following condition holds:

(5.1) $\qquad \exists\, \delta > 0,\ such\ that\ p_\delta^n := \dfrac{\#\Big(\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap F_n\Big)}{\#F_n} \to p_\delta \in [0, 1).$

This Proposition gives a condition for the continuity of the slope functional $LS(\cdot)$ when $\{F_n n \geq 1\}$ and F are bounded sets in $\mathcal{F}_{\text{finite or line}}$. The next example shows the necessity of condition (5.1), which prevents a set of outlier points from skewing the slope of the least squares line.

Example 5.1. For $n \geq 1$, define the sets:

$$F_n = \{(\frac{i}{n}, 0), -n \leq i \leq n; (\frac{1}{n}(1 + \frac{j}{2^n}), \frac{1}{n}(1 + \frac{j}{2^n})), 0 \leq j \leq 2^n\} \quad \text{and} \quad F = [-1, 1] \times \{0\}.$$

We develop features about this example.

(1) For the cardinality of F_n we have

$$\#F_n = k_n = 2^n + 2n + 2.$$

(2) We have $F_n \to F$ in $\mathcal{K}_2$. As before, denote the Hausdorff distance between two closed sets in $\mathcal{K}_2$ by $D(\cdot, \cdot)$ and we have $D(F_n, F) < 3/n \to 0$ as $n \to \infty$.

(3) Condition (5.1) is *not* satisfied. To see this pick any $n \geq 1$ and observe

$$\bar{x}_n = \bar{y}_n = \frac{3(2^n + 1)}{2n(2^n + 2n + 2)} = \frac{3(2^n + 1)}{2nk_n} \sim \frac{3}{2n}.$$

Fix $\delta > 0$. For all n so large that $\delta > 1/(2n)$ we have

$$\frac{\#\left(\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap F_n\right)}{\#F_n} \geq \frac{2^n + 1}{2^n + 2n + 2} \to 1, \quad (n \to \infty).$$

Obviously for this example, $m = LS(F) = 0$. However, if m_n denotes the slope of the least squares line through F_n then we show that $m_n \to 1 \neq 0 = m$. To see this, observe that conventional wisdom yields,

(5.2)
$$m_n = \frac{\displaystyle\sum_{(x_i(n), y_i(n)) \in F_n} (y_i(n) - \bar{y})(x_i(n) - \bar{x})}{\displaystyle\sum_{(x_i(n), y_i(n)) \in F_n} (x_i(n) - \bar{x})^2}.$$

For the numerator we have,

$$\sum_{(x_i(n), y_i(n)) \in F_n} (y_i(n) - \bar{y}_n)(x_i(n) - \bar{x}_n) = \sum_{(x_i(n), y_i(n)) \in F_n} y_i(n)x_i(n) - k_n \bar{y}_n \bar{x}_n$$

$$= \sum_{j=0}^{2^n} \frac{1}{n^2}(1 + \frac{j}{2^n})^2 - k_n\left(\frac{3(2^n + 1)}{2nk_n}\right)^2 = \frac{1}{n^2}\left(\sum_{j=0}^{2^n}\left(1 + \frac{2j}{2^n} + \frac{j^2}{2^{2n}}\right) - \frac{9}{4k_n}(2^n + 1)^2\right)$$

$$= \frac{1}{n^2}\left(2 \cdot (2^n + 1) + \frac{1}{2^{2n}}\sum_{j=0}^{2^n} j^2 - \frac{9}{4k_n}(2^n + 1)^2\right)$$

and using the identity $\sum_{j=1}^{N} j^2 = N(N+1)(N+\frac{1}{2})/3 = N(N+1)(2N+1)/6$, we get the above equal to

$$= \frac{1}{n^2}\left(2 \cdot (2^n + 1) + \frac{1}{2^{2n}}\frac{2^n(2^n + 1)(2^n + \frac{1}{2})}{3} - \frac{9}{4k_n}(2^n + 1)^2\right)$$

$$= \frac{2^n + 1}{n^2}\left(2 + \frac{2^n + \frac{1}{2}}{3 \cdot 2^n} - \frac{9}{4k_n}(2^n + 1)\right) \sim \frac{k_n}{12n^2}.$$

For the denominator, we use the calculation already done for the numerator:

$$\sum_{(x_i(n), y_i(n)) \in F_n} (x_i(n) - \bar{x}_n)^2 = \sum_{(x_i(n), y_i(n)) \in F_n} x_i(n)^2 - k_n(\bar{x}_n)^2$$

$$= \sum_{i=-n}^{n} (\frac{j}{n})^2 + \sum_{j=0}^{2^n} \frac{1}{n^2}(1 + \frac{j}{2^n})^2 - k_n\left(\frac{3(2^n + 1)}{2nk_n}\right)^2$$

$$= \sum_{i=-n}^{n} (\frac{j}{n})^2 + \sum_{(x_i(n), y_i(n)) \in F_n} y_i(n)x_i(n) - k_n \bar{y}_n \bar{x}_n$$

$$= \frac{2n(n+1)(2n+1)}{6n^2} + \frac{k_n}{12n^2} + o(\frac{k_n}{12n^2})$$

$$= O(n) + \frac{k_n}{12n^2} + o(\frac{k_n}{12n^2}) \sim \frac{k_n}{12n^2}.$$

Combining the asymptotic forms for numerator and denominator with (5.2) yields

$$m_n \sim \frac{k_n/12n^2}{k_n/12n^2} \sim 1, \quad (n \to \infty),$$

so $m_n \to 1 \neq 0 = m$, as claimed. $\square$

Proof of Proposition 5.1. For $(x_i(n), y_i(n)) \in F_n$, we can write

(5.3) $$y_i(n) = mx_i(n) + z_i(n) \qquad 1 \leq i \leq k_n$$

We want to show that $m_n = LS(F_n) \to m = LS(F)$, as $n \to \infty$. Fix $\epsilon > 0$. We will provide N such that for $n > N$, we have $|m_n - m| < \epsilon$.

First of all, condition (5.1) allows us to fix $\delta > 0$ such that

$$p_\delta^n := p_n = \frac{\#\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap F_n}{\#F_n} \to p < 1.$$

Choose N_1 such that for $n > N_1$, we have $p_n < \frac{1+p}{2}$ or equivalently that $1 - p_n > \frac{1-p}{2}$. For $\eta > 0$ and $F \in \mathcal{K}_2$, recall the definition of the η-swelling of F:

(5.4) $$F^\eta = \{x : d(x, y) < \eta \text{ for some } y \in F\}.$$

Since $D(F_n, F) \to 0$ in $\mathcal{K}_2$, we can choose N_2 such that for all $n > N_2$ we have $F_n \subset F^{\epsilon_1}$ where

$$\epsilon_1 := \frac{2\delta\epsilon(1-p)}{4\sqrt{1+m^2}(2+2m+\epsilon(1-p))} = \delta_1\epsilon\frac{(1-p)}{4\sqrt{1+m^2}}$$

and we have set

$$\delta_1 := \frac{\delta}{1 + m + \frac{1}{2}\epsilon(1-p)} < \delta.$$

The choice of δ_1 is designed to ensure that if for some $(x_i(n), y_i(n))$, we have $|x_i(n) - \bar{x}_n| < \delta_1$, then

$$(x_i(n), y_i(n)) \in (\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta).$$

This follows because

$$|x_i(n) - \bar{x}_n| \vee |y_i(n) - \bar{y}_n| < \delta_1 + m\delta_1 + 2\epsilon_1\sqrt{1+m^2}.$$

See Figure 2; from the definition of ϵ_1 we have this equal to
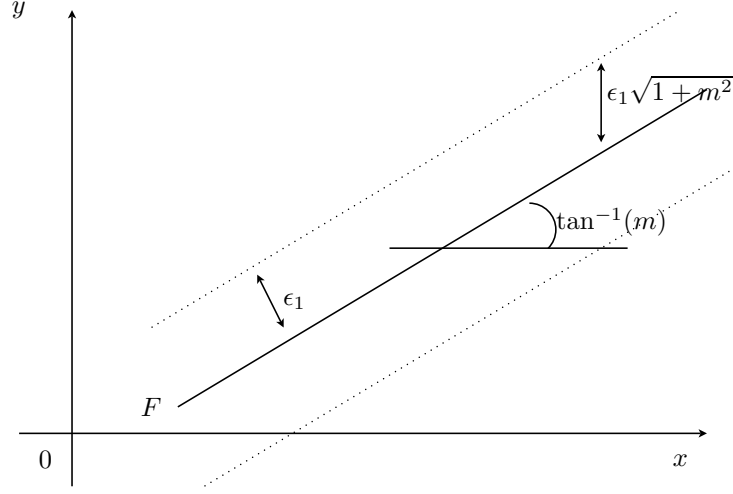
(5.5) $$= \delta_1 + m\delta_1 + 2\frac{\delta_1\epsilon(1-p)}{4\sqrt{1+m^2}}\sqrt{1+m^2} = \delta_1(1 + m + \frac{\epsilon(1-p)}{2}) = \delta.$$

Let $N = N_1 \vee N_2$ and restrict attention to $n > N$. Since $F_n \subset F^{\epsilon_1}$, we have for all $1 \leq i \leq k_n$ that $(x_i(n), y_i(n)) \in F^{\epsilon_1}$. By convexity of F^{ϵ_1}, $(\bar{x}, \bar{y}) \in F^{\epsilon_1}$. Therefore, referring to Figure 1, we have

(5.6) $$\begin{aligned} |z_i(n) - \bar{z}_n| &\leq |y_i(n) - mx_i(n)| + |\bar{y}_n - m\bar{x}_n| \\ &\leq \epsilon_1\sqrt{1+m^2} + \epsilon_1\sqrt{1+m^2} = 2\epsilon_1\sqrt{1+m^2}. \end{aligned}$$

Using the representation (5.3) we get,

(5.7) $$m_n = \frac{\sum_{i=1}^{k_n}(y_i(n) - \bar{y}_n)(x_i(n) - \bar{x}_n)}{\sum_{i=1}^{k_n}(x_i(n) - \bar{x}_n)^2} = m + \frac{\sum_{i=1}^{k_n}(z_i(n) - \bar{z}_n)(x_i(n) - \bar{x}_n)}{\sum_{i=1}^{k_n}(x_i(n) - \bar{x}_n)^2}.$$

FIGURE 1. The geometry of the neighborhood of the line F.

Therefore,

$$|m_n - m| = \left| \frac{\sum\limits_{i=1}^{k_n} (z_i(n) - \bar{z}_n)(x_i(n) - \bar{x}_n)}{\sum\limits_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2} \right| \leq \frac{\sum\limits_{i=1}^{k_n} |z_i(n) - \bar{z}_n||x_i(n) - \bar{x}_n|}{\sum\limits_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2} \leq 2\epsilon_1 \sqrt{1+m^2} \frac{\sum\limits_{i=1}^{k_n} |x_i(n) - \bar{x}_n|}{\sum\limits_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2}$$

where the last inequality follows from (5.6).

For convenience, define the following notation:

$$|S(x)|_{<\rho} := \sum_{|x_i(n)-\bar{x}_n|<\rho} |x_i(n) - \bar{x}_n|, \qquad |S(x)|_{\geq\rho} := \sum_{|x_i(n)-\bar{x}_n|\geq\rho} |x_i(n) - \bar{x}_n|,$$

$$S^2(x)_{<\rho} := \sum_{|x_i(n)-\bar{x}_n|<\rho} (x_i(n) - \bar{x}_n)^2, \qquad S^2(x)_{\geq\rho} := \sum_{|x_i(n)-\bar{x}_n|\geq\rho} (x_i(n) - \bar{x}_n)^2,$$

$$B\big((x,y),\delta\big) := (x-\delta, x+\delta) \times (y-\delta, y+\delta).$$

Therefore

$$\frac{\sum\limits_{i=1}^{k_n} |x_i(n) - \bar{x}_n|}{\sum\limits_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2} = \frac{|S(x)|_{<\delta_1} + |S(x)|_{\geq\delta_1}}{S^2(x)_{<\delta_1} + S^2(x)_{\geq\delta_1}} = \frac{(|S(x)|_{<\delta_1} + |S(x)|_{\geq\delta_1})/S^2(x)_{\geq\delta_1}}{(S^2(x)_{<\delta_1}/S^2(x)_{\geq\delta_1} + 1)}$$

$$\leq \frac{(|S(x)|_{<\delta_1} + |S(x)|_{\geq\delta_1})}{S^2(x)_{\geq\delta_1}} \leq \frac{1}{\delta_1}\left(\frac{|S(x/\delta_1)|_{<1}}{S^2(x/\delta_1)_{\geq1}} + 1 \right)$$

$$\leq \frac{1}{\delta_1}\left(\frac{\#\{(x_i(n), y_i(n)) \in F_n : |x_i(n) - \bar{x}_n| < \delta_1\}}{\#\{(x_i(n), y_i(n)) \in F_n : |x_i(n) - \bar{x}_n| \geq \delta_1\}} + 1 \right)$$

$$\leq \frac{1}{\delta_1}\left(\frac{\#\{(x_i(n), y_i(n)) \in F_n : (x_i(n), y_i(n)) \in B((\bar{x}_n, \bar{y}_n), \delta)\}}{\#\{(x_i(n), y_i(n)) \in F_n : (x_i(n), y_i(n)) \notin B((\bar{x}_n, \bar{y}_n), \delta)\}} + 1 \right)$$
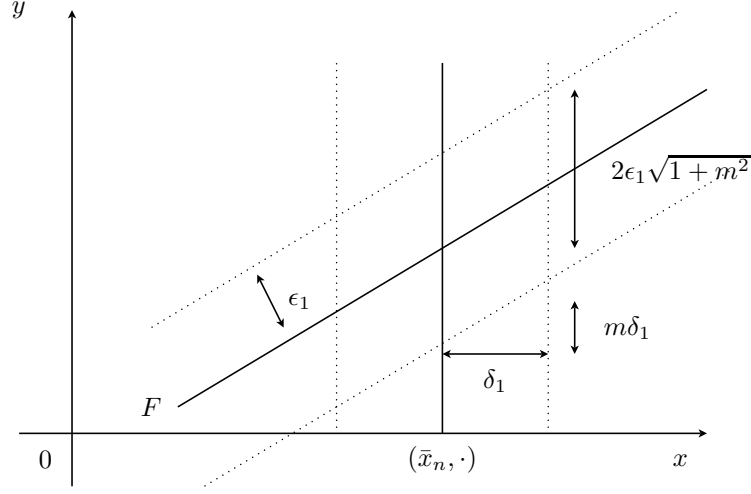
FIGURE 2. The geometry of a neighborhood of the line; support for (5.5).

The choice of δ_1 justifies the previous step by (5.5). The previous expression is bounded by

$$\leq \frac{1}{\delta_1}(\frac{p_n}{1-p_n}+1) \leq \frac{1}{\delta_1}(\frac{1+p}{1-p}+1) = \frac{2}{\delta_1(1-p)},$$

and we recall $p < 1$.

Consequently

$$|m_n - m| = 2\epsilon_1\sqrt{1+m^2} \frac{\displaystyle\sum_{i=1}^{k_n}|x_i(n)-\bar{x}_n|}{\displaystyle\sum_{i=1}^{k_n}(x_i(n)-\bar{x}_n)^2} \leq 2\epsilon_1\sqrt{1+m^2} \times \frac{2}{\delta_1(1-p)}$$

$$= 2\epsilon\delta_1 \frac{1-p}{4\sqrt{1+m^2}}\sqrt{1+m^2} \times \frac{2}{\delta_1(1-p)} = \epsilon$$

This completes the proof that $m_n \to m$ under condition (5.1). $\qquad\square$

Corollary 5.2. If $\bar{x}_n \to \mu_x < \infty$ and $\bar{y}_n \to \mu_y < \infty$, as $n \to \infty$, then Proposition 5.1 holds if we replace $(\bar{x}_n, \bar{y}_n)$ in (5.1) by (μ_x, μ_y).

Proof. In place of condition (5.1) we are assuming

(5.8) $\exists\, \delta > 0$ such that $p_\delta^n = \dfrac{\#\{(\mu_x - \delta, \mu_x + \delta) \times (\mu_y - \delta, \mu_y + \delta)\} \bigcap F_n}{\#F_n} \to p_\delta \in [0,1).$

Let us fix $\delta > 0$ such that

$$p_n^* := \frac{\#\{(\mu_x - 2\delta, \mu_x + 2\delta) \times (\mu_y - 2\delta, \mu_y + 2\delta)\} \bigcap F_n}{\#F_n} \to p \in [0,1).$$

Since $\bar{x}_n \to \mu_x < \infty$ and $\bar{y}_n \to \mu_y < \infty$, there exists N^* such that $n > N^*$ implies that $(\bar{x}_n, \bar{y}_n) \in (\mu_x - \delta, \mu_x + \delta) \times (\mu_y - \delta, \mu_y + \delta)$. Hence for $n > N^*$

$$p_n := \frac{\#\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap F_n}{\#F_n}$$

$$\leq \frac{\#\{(\mu_x - 2\delta, \mu_x + 2\delta) \times (\mu_y - 2\delta, \mu_y + 2\delta)\} \bigcap F_n}{\#F_n}$$

$$= p_n^* \to p \in [0,1).$$

Now choose $N_1 \geq N^*$ such that for all $n > N_1$, we have $p_n < \frac{1+p}{2}$. This also means that $1 - p_n > \frac{1-p}{2}$. The rest of the proof is the same as that of Proposition 5.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6. Slope of the LS line as a tail index estimator

For heavy tailed distributions, the slope of the least squares line through the QQ plot made by the upper k_n largest order statistics is a consistent estimator of $1/\alpha$. See Beirlant et al. (1996), Kratz and Resnick (1996) and Resnick (2006, Section 4.6). We connect the ideas of the previous section with this result.

Proposition 6.1. *Consider non-negative random variables $X_1, X_2, \ldots, X_n$ which are iid with common distribution F where $\bar{F} \in RV_{-\alpha}$ and $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$ are the order statistics in decreasing order. The sets $\mathcal{S}_n$ and $\mathcal{T}_n$ were defined in Proposition 4.1 where we proved $d_{\mathcal{F}}(\mathcal{S}_n, \mathcal{T}_n) \overset{P}{\to} 0$, assuming $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. For convenience we defined $\mathcal{S}'_n = \mathcal{S}_n + a_n$ and $\mathcal{T} = \mathcal{T}_n + a_n$ where a_n was a random point. Write*

$$\mathcal{S}'_n = \{(-\log \frac{j}{k_n}, \log \frac{X_{(j)}}{X_{(k)}}; j = 1, \ldots, k_n\} = \{(x_j(n), y_j(n)); j = 1, \ldots, k_n\} \ (\text{ say}) \text{ and } \mathcal{T} = \{(x, \frac{x}{\alpha}); x \geq 0\}.$$

Then,

(6.1) $$LS(\mathcal{S}'_n) = LS(\mathcal{S}_n) \overset{P}{\to} \frac{1}{\alpha} = LS(\mathcal{T}_n) = LS(\mathcal{T}),$$

as $k := k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

The result is believable based on the fact that $d_{\mathcal{F}}(\mathcal{S}_n, \mathcal{T}_n) \overset{P}{\to} 0$. However, since neither $\mathcal{T}_n$ nor $\mathcal{T}$ are $\mathcal{K}_2$ sets, some sort of truncation to compact regions of $\mathbb{R}^2$ is necessary in order to capitalize on Proposition 5.1. For some integer $M > 2$, define

$$K_M = [0, M] \times [0, \frac{2M}{\alpha}],$$

and let

$$\mathcal{S}'^M_n = \mathcal{S}'_n \cap K_M \quad \text{and} \quad \mathcal{T}^M = \mathcal{T} \cap K_M.$$

Proof. Some preliminary observations. Clearly, $LS(\mathcal{S}_n) = LS(\mathcal{S}'_n + a_n) = LS(\mathcal{S}'_n)$ and with $x_j(n), y_j(n)$ defined in the statement of the Proposition,

$$LS(\mathcal{S}'_n) = \frac{\bar{S}_{XY} - \bar{S}_X \bar{S}_Y}{\bar{S}_{XX} - (\bar{S}_X)^2},$$

where, as usual,

$$\bar{S}_X = \frac{1}{k_n} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'_n} x_j(n), \qquad\qquad \bar{S}_Y = \frac{1}{k_n} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'_n} y_j(n),$$

$$\bar{S}_{XY} = \frac{1}{k_n} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'_n} x_j(n) y_j(n), \qquad\qquad \bar{S}_{XX} = \frac{1}{k_n} \sum_{(x_j(n), x_j(n)) \in \mathcal{S}'_n} (x_j(n))^2.$$

We need similar quantities $\bar{S}^M_X, \bar{S}^M_Y, \bar{S}^M_{XY}$ corresponding to averages of points restricted to K_M, so for instance

$$\bar{S}^M_X = \frac{1}{k^M} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'^M_n} x_j(n)$$

and $k^M = \#\mathcal{S}'^M_n$. A simple calculation given in Resnick (2006, page 109) yields as $k \to \infty$,

(6.2) $$\bar{S}_X = \frac{1}{k} \sum_{i=1}^{k} (-\log \frac{i}{k}) \sim \int_0^1 (-\log x) dx = 1, \quad \bar{S}_{XX} = \frac{1}{k} \sum_{i=1}^{k} (-\log \frac{i}{k})^2 \sim \int_0^1 (-\log x)^2 dx = 2,$$

while for $\bar{S}_Y$ we have

$$(6.3) \qquad \bar{S}_Y = \frac{1}{k} \sum_{i=1}^{k} (-\log \frac{X_{(i)}}{X_{(k)}}) \xrightarrow{P} \frac{1}{\alpha}$$

since $\bar{S}_Y$ is the Hill estimator and is consistent for $1/\alpha$ (Csörgő et al., 1985, Mason, 1982, Mason and Turova, 1994, Resnick, 2006).

We need the corresponding limits for $\bar{S}_X^M, \bar{S}_{XX}^M, \bar{S}_Y^M$. These calculations and subsequent calculations are simplified by the following facts:

(1) The ratios of order statistics process converges, as $k \to \infty$, $k/n \to 0$,

$$(6.4) \qquad \frac{X_{(\lceil kt \rceil)}}{X_{(k)}} \xrightarrow{P} t^{-1/\alpha},$$

in $D_l(0, \infty]$ (Resnick, 2006, page 82).

(2) Define the random measure

$$\hat{\nu}_n(\cdot) = \frac{1}{k} \sum_{i=1}^{n} \epsilon_{X_{(i)}/X_{(k)}}(\cdot)$$

on $(0, \infty]$, which puts mass $1/k$ at the points $\{X_{(i)}/X_{(k)}, 1 \le i \le n\}$. Then

$$(6.5) \qquad \hat{\nu}_n \xrightarrow{P} \nu_\alpha,$$

in the space of Radon measures on $(0, \infty]$, where $\nu_\alpha(x, \infty] = x^{-\alpha}$, $x > 0$ (Resnick, 2006, page 83).

(3) The number of points k^M in $\mathcal{S}_n'^M$ satisfies, as $n \to \infty$, $k \to \infty$, $k/n \to 0$,

$$(6.6) \qquad k^M/k \xrightarrow{P} 1 - e^{-M}.$$

To see this, observe

$$k^M/k = \frac{1}{k} \#\{j \le k : k \ge j \ge ke^{-M} \text{ and } \frac{X_{(j)}}{X_{(k)}} \le e^{2M/\alpha}\}$$

$$= \frac{1}{k} \#\{j \le k : 1 \le \frac{X_{(j)}}{X_{(k)}} \le \frac{X_{(\lceil ke^{-M} \rceil)}}{X_{(k)}} \wedge e^{2M/\alpha}\}$$

$$= \hat{\nu}_n \left(1, \frac{X_{(\lceil ke^{-M} \rceil)}}{X_{(k)}} \wedge e^{2M/\alpha}\right]$$

$$\xrightarrow{P} 1 - \left((e^{-M})^{-1/\alpha} \wedge e^{2M/\alpha}\right)^{-\alpha} = 1 - e^{-M}.$$

We continue using these three facts. For $\bar{S}_X^M$ we have

$$\bar{S}_x^M = \frac{1}{k^M} \sum_{\left(x_i(n), y_i(n)\right) \in \mathcal{S}_n'^M} x_i(n) = \frac{1}{k^M} \sum_{\substack{j:k \ge j \ge ke^{-M} \\ 0 < \log X_{(j)}/X_{(k)} \le 2M/\alpha}} -\log \frac{j}{k}.$$

Set

$$\left(\bar{S}_X^M\right)^* := \frac{1}{k^M} \sum_{j:k \ge j \ge ke^{-M}} -\log \frac{j}{k} = \frac{k}{k^M} \frac{1}{k} \sum_{j:k \ge j \ge ke^{-M}} -\log \frac{j}{k}$$

$$\sim \frac{1}{1 - e^{-M}} \int_{e^{-M}}^{1} -\log x \, dx = \frac{1}{1 - e^{-M}} \int_0^M y e^{-y} dy$$

$$=: 1 + \epsilon_X(M),$$

where $\epsilon_X(M) \to 0$ as $M \to \infty$. Also, $\bar{S}_X^M$ and $\left(\bar{S}_X^M\right)^*$ are close asymptotically since

$$P[\bar{S}_X^M \neq \left(\bar{S}_X^M\right)^*] = P\Big\{ \bigcup_{k \geq j \geq k^{-M}} [\log \frac{X_{(j)}}{X_{(k)}} > 2M/\alpha] \Big\}$$

$$= P[\log \frac{X_{(\lceil ke^{-M} \rceil)}}{X_{(k)}} > 2M/\alpha] \to 0,$$

since

$$\frac{X_{(\lceil ke^{-M} \rceil)}}{X_{(k)}} \xrightarrow{P} e^{M/\alpha} < e^{2M/\alpha}.$$

We conclude

(6.7) $$\bar{S}_X^M \xrightarrow{P} 1 + \epsilon_X(M) := \mu_X^M,$$

with $\epsilon_X(M) \to 0$ as $M \to \infty$, and in a similar way we can derive that

(6.8) $$\bar{S}_{XX}^M \xrightarrow{P} 2 + \epsilon_{XX}(M),$$

where $\epsilon_{XX}(M) \to 0$ as $M \to \infty$. For $\bar{S}_Y^M$ we have

$$\bar{S}_Y^M = \frac{1}{k^M} \sum_{\substack{j:k \geq j \geq ke^{-M} \\ 0 < \log X_{(j)}/X_{(k)} \leq 2\alpha^{-1}M}} \log \frac{X_{(j)}}{X_{(k)}}$$

$$= \frac{1}{k^M} \sum_{j:0 < \log X_{(j)}/X_{(k)} \leq 2\alpha^{-1}M \wedge \log X_{(\lceil ke^{-M} \rceil)}/X_{(j)}} \log \frac{X_{(j)}}{X_{(k)}}$$

$$= \frac{k}{k^M} \int_1^{2\alpha^{-1}M \wedge \log X_{(\lceil ke^{-M} \rceil)}/X_{(j)}} \log y \, \hat{\nu}_n(dy)$$

$$\xrightarrow{P} \frac{1}{1 - e^{-M}} \int_1^{2\alpha^{-1}M \wedge \alpha^{-1}M} \log y \, \nu_\alpha(dy)$$

$$= \frac{1}{1 - e^{-M}} \int_0^{M/\alpha} s e^{-\alpha s} ds =: \mu_Y^M,$$

where $\mu_Y^M \to \frac{1}{\alpha}$ as $M \to \infty$. We conclude

(6.9) $$\bar{S}_Y^M \xrightarrow{P} \mu_Y^M.$$

To prove (6.1), we follow the following outline of steps.

- Step 1: Prove $\mathcal{S}_n'^M \xrightarrow{P} \mathcal{T}^M$.
- Step 2: Verify that Corollary 5.2 is applicable by showing that the analogue of (5.1) holds. This permits the conclusion that

$$LS(\mathcal{S}_n'^M) \xrightarrow{P} 1/\alpha.$$

Coupled with (6.7), (6.8) and (6.9), this yields

(6.10) $$\bar{S}_{XY}^M = \frac{2}{\alpha} + \epsilon_{XY}(M) + o_p(1),$$

where $\lim_{M \to \infty} \epsilon_{XY}(M) = 0$ and $o_p(1) \xrightarrow{P} 0$ as $n \to \infty$.
- Step 3: Compare $\bar{S}_{XY}$ and $\bar{S}_{XY}^M$ and Check that

(6.11) $$\lim_{M \to \infty} \limsup_{n \to \infty} P[|\bar{S}_{XY}^M - \bar{S}_{XY}| > \eta] = 0, \quad \forall \eta > 0.$$

This gives $\bar{S}_{XY} \xrightarrow{P} 2/\alpha$ which coupled with (6.2) and (6.3) implies (6.1).

We may check Step 1 using a very minor modification of Lemma 2.3, following the pattern of proof used for Proposition 4.1. For Step 2, the challenge is to verify condition (5.1) holds and we defer this to the end of the proof. Thus we turn to Step 3.

First of all, we observe that $\bar{S}_{XY}^M$ and $\bar{S}_{XY}$ average, respectively k^M and k terms but there is no need to differentiate: For any $\eta > 0$,

$$P\Big[\Big|\frac{1}{k^M}\sum_{(x_j(n),y_j(n))\in\mathcal{S}_n'^M}x_i(n)y_i(n) - \frac{1}{k}\sum_{(x_j(n),y_j(n))\in\mathcal{S}_n'^M}x_i(n)y_i(n)\Big| > \eta\Big]$$

$$= P\Big[\Big|\frac{1}{k^M} - \frac{1}{k}\Big|\sum_{(x_j(n),y_j(n))\in\mathcal{S}_n'^M}x_i(n)y_i(n) > \eta\Big]$$

and dividing the sum by k^M yields

$$= P\Big[\bar{S}_{XY}^M\Big|1 - \frac{k^M}{k}\Big| > \eta\Big].$$

Since $\bar{S}_{XY}^M$ is convergent in probability, it is stochastically bounded and since, as $n \to \infty$,

$$\Big|1 - \frac{k^M}{k}\Big| \xrightarrow{P} 1 - (1 - e^{-M}) = e^{-M} \xrightarrow{M\to\infty} 0,$$

we conclude

(6.12) $$\lim_{M\to\infty}\limsup_{n\to\infty} P\Big[\Big|\frac{1}{k^M}\sum_{(x_j(n),y_j(n))\in\mathcal{S}_n'^M}x_i(n)y_i(n) - \frac{1}{k}\sum_{(x_j(n),y_j(n))\in\mathcal{S}_n'^M}x_i(n)y_i(n)\Big| > \eta\Big] = 0.$$

Next observe for $\eta > 0$,

$$P\Big[\Big|\frac{1}{k}\sum_{(x_j(n),y_j(n))\in\mathcal{S}_n'^M}x_j(n)y_j(n) - \frac{1}{k}\sum_{k\geq j\geq ke^{-M}}x_j(n)y_j(n)\Big| > \eta\Big] \leq P\Big\{\bigcup_{k\geq j\geq ke^{-M}}[\frac{X_{(j)}}{X_{(k)}} > e^{2M/\alpha}]\Big\}$$

(6.13) $$\leq P[\frac{X_{(\lceil ke^{-M}\rceil)}}{X_{(k)}} > e^{2M/\alpha}] \to 0, \quad (n \to \infty).$$

Note that by the Cauchy-Schwartz inequality,

$$\Big(|\bar{S}_{XY} - \frac{1}{k}\sum_{k\geq j\geq ke^{-M}}x_j(n)y_j(n)|\Big)^2 \leq \Big(\frac{1}{k}\sum_{1\leq j\leq ke^{-M}}x_j(n)y_j(n)\Big)^2 \leq \frac{1}{k}\sum_{1\leq j\leq ke^{-M}}x_j(n)^2 \cdot \frac{1}{k}\sum_{1\leq j\leq ke^{-M}}y_j(n)^2.$$

Furthermore

$$\frac{1}{k}\sum_{1\leq j\leq ke^{-M}}y_j(n)^2 = \int_{\log X_{(\lceil ke^{-M}\rceil)}/X_{(k)}}^{\infty}(\log y)^2\hat{\nu}_n(dy)$$

and using (6.4), we have for some $c > 0$, all large n and some M that the above is bounded by

(6.14) $$\int_{cM}^{\infty}(\log y)^2\hat{\nu}_n(dy) + o_p(1).$$

Assessing (6.12), (6.13) and (6.14), we see that (6.11) will be proved if we show

(6.15) $$\lim_{M\to\infty}\limsup_{n\to\infty} P\Big[\int_M^{\infty}(\log y)^2\hat{\nu}_n(dy) > \eta\Big] = 0, \quad (\forall\eta > 0).$$

This treatment is similar to the stochastic version of Karamata's theorem (Feigin and Resnick (1997), Resnick (2006, page 207). For $0 < \zeta < 1 \wedge \alpha$ and large M, the integrand $(\log y)^2$ is dominated by y^ζ. Bound the integral by

$$\int_M^{\infty}\hat{\nu}_n(y,\infty]\,\zeta y^{\zeta-1}dy + M^\zeta\hat{\nu}_n(M,\infty].$$

If we let first $n \to \infty$ and then $M \to \infty$, for the second piece we have

$$M^\zeta \hat{\nu}_n(M, \infty] \xrightarrow{P} M^\zeta \nu_\alpha(M, \infty] = M^{\zeta-\alpha} \to 0.$$

Now we deal with the integral. Set $b(t) = (1/(1-F))^\leftarrow(t)$ so that $X_{(k)}/b(n/k) \xrightarrow{P} 1$ (Resnick, 2006, page 81). For $\gamma > 0$,

$$P[\int_M^\infty \hat{\nu}_n(y, \infty] \, \zeta y^{\zeta-1} dy > \eta] = P[\int_M^\infty \hat{\nu}_n(y, \infty] \, \zeta y^{\zeta-1} dy > \eta, 1 - \gamma < X_{(k)}/b(n/k) < 1 + \gamma] + o(1)$$

$$\leq P[\int_M^\infty \frac{1}{k} \sum_{i=1}^n \epsilon_{X_i/b(n/k)}((1-\gamma)y, \infty] \zeta y^{\zeta-1} dy > \eta] + o(1).$$

Ignore the term $o(1)$. Markov's inequality gives a bound

$$\leq (const) \int_M^\infty E\Big(\frac{1}{k} \sum_{i=1}^n P[X_i \geq b(n/k)(1-\gamma)y]\Big) \zeta y^{\zeta-1} dy$$

$$= (const) \int_M^\infty \frac{n}{k} \bar{F}(b(n/k)(1-\gamma)y]) \zeta y^{\zeta-1} dy.$$

and applying Karamata's theorem (Bingham et al., 1987, de Haan, 1970, Geluk and de Haan, 1987, Resnick, 2006), we have as $n \to \infty$ that this converges to

$$= (const) \int_M^\infty ((1-\gamma)y)^{-\alpha} \zeta y^{\zeta-1} dy \xrightarrow{M \to \infty} 0,$$

as required. This finishes Step 3 and completes the proof modulo the verification that (5.8) can be proven for this problem.

The remaining task of checking (5.8) proceeds as follows. Recall μ_X^M and μ_Y^M from (6.7) and (6.9). Fix M. Then for p_δ^n in (5.8), we have

$$\frac{1}{k^M} \#\{j : \mu_X^M - \delta < -\log \frac{j}{k} < \mu_X^M + \delta, 0 < -\log \frac{j}{k} \leq M; \mu_Y^M - \delta < \log \frac{X_{(j)}}{X_{(k)}} < \mu_Y^M + \delta, 0 \leq \log \frac{X_{(j)}}{X_{(k)}} \leq \frac{2M}{\alpha}\}.$$

Since $\mu_X^M \approx 1$ and $\mu_Y^M \approx 1/\alpha$, we get for large M

$$p_\delta^n := \frac{1}{k^M} \#\Big\{j : \mu_X^M - \delta < -\log \frac{j}{k} < \mu_X^M + \delta; \ \mu_Y^M - \delta < \log \frac{X_{(j)}}{X_{(k)}} < \mu_Y^M + \delta\Big\}$$

$$= \frac{1}{k^M} \#\Big\{j : \frac{X_{(\lceil k \exp\{-(\mu_X^M - \delta)\}\rceil)}}{X_{(k)}} \vee e^{\mu_Y^M - \delta} < \frac{X_{(j)}}{X_{(k)}} < \frac{X_{(\lceil k \exp\{-(\mu_X^M + \delta)\}\rceil)}}{X_{(k)}} \wedge e^{\mu_Y^M + \delta}\Big\}$$

$$= \frac{k}{k^M} \hat{\nu}_n\Big(\frac{X_{(\lceil k \exp\{-(\mu_X^M - \delta)\}\rceil)}}{X_{(k)}} \vee e^{\mu_Y^M - \delta}, \frac{X_{(\lceil k \exp\{-(\mu_X^M + \delta)\}\rceil)}}{X_{(k)}} \wedge e^{\mu_Y^M + \delta}\Big).$$

Apply (6.4) and (6.5) and we find

$$p_\delta^n \xrightarrow{P} \frac{1}{1 - e^{-M}} \nu_\alpha \Big(e^{(\mu_X^M - \delta)/\alpha} \vee e^{\mu_Y^M - \delta}, e^{(\mu_X^M + \delta)/\alpha} \wedge e^{\mu_Y^M + \delta}\Big).$$

Since $\mu_X^M \approx 1$ and $\mu_Y^M \approx 1/\alpha$, by picking M large and δ small, the right side can be made to be less than 1. This completes the proof. $\square$

References

G. Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993. ISBN 0-7923-2531-1.

J. Beirlant, P. Vynckier, and J. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, 91(436):1659–1667, 1996. ISSN 0162-1459.

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

S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates for the tail index of a distribution. *Ann. Statist.*, 13: 1050–1077, 1985.

L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematisch Centrum Amsterdam, 1970.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer-Verlag, New York, 2006.

P. Feigin and S.I. Resnick. Linear programming estimators and bootstrapping for heavy tailed phenomena. *Adv. in Appl. Probab.*, 29:759–805, 1997.

J. Flachsmeyer. Verschiedene topologisierungen im Raum der abgeschlossenen Mengen. *Math. Nachr.*, 26:321–337, 1963/1964. ISSN 0025-584X.

J. L. Geluk and L. de Haan. *Regular Variation, Extensions and Tauberian Theorems*, volume 40 of *CWI Tract*. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 1987. ISBN 90-6196-324-9.

M. Kratz and S.I. Resnick. The qq–estimator and heavy tails. *Stochastic Models*, 12:699–724, 1996.

D. Mason. Laws of large numbers for sums of extreme values. *Ann. Probab.*, 10:754–764, 1982.

D. Mason and T. Turova. Weak convergence of the Hill estimator process. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme Value Theory and Applications*, pages 419–432. Kluwer Academic Publishers, Dordrecht, Holland, 1994.

G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, New York-London-Sydney, 1975. With a foreword by G.S. Watson, Wiley Series in Probability and Mathematical Statistics.

I. Molchanov. *Theory of Random Sets*. Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2005. ISBN 978-185223-892-3; 1-85233-892-X.

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.

S.I. Resnick. *A Probability Path*. Birkhäuser, Boston, 1999.

S.I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2006. ISBN: 0-387-24272-4.

E. Seneta. *Regularly Varying Functions*. Springer-Verlag, New York, 1976. Lecture Notes in Mathematics, 508.

R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons Inc., New York, 1980. ISBN 0-471-02403-1.

N. V. Smirnov. Limit distributions for the terms of a variational series. *Trudy Mat. Inst. Steklov.*, 25:60, 1949. ISSN 0371-9685.

W. Vervaat. Random upper semicontinuous functions and extremal processes. In W. Vervaat and H. Holwerda, editors, *Probability and Lattices*, volume 110 of *CWI Tract*, pages 1–56. Math. Centrum Wisk. Inform., Amsterdam, 1997.

Bikramjit Das, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853
  E-mail address: bd72@cornell.edu

Sidney Resnick, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853
  E-mail address: sir1@cornell.edu