SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-3801

TECHNICAL REPORT NO. 1346

September 2002

THE EXTREMAL DEPENDENCE MEASURE AND ASYMPTOTIC INDEPENDENCE

by

Sidney Resnick

THE EXTREMAL DEPENDENCE MEASURE AND ASYMPTOTIC INDEPENDENCE

SIDNEY RESNICK

ABSTRACT. Extremal dependence analysis assesses the tendency of large values of components of a random vector to occur simultaneously. This kind of dependence information can be qualitatively different than what is given by correlation which averages over the total body of the joint distribution. Also, correlation may be completely inappropriate for heavy tailed data. We study the *extremal dependence measure* (EDM), a measure of the tendency of large values of components of a random vector to occur simultaneously and show consistency and asymptotic normality properties for the standard case of multivariate regular variation.

1. INTRODUCTION

Extremal dependence analysis assesses the tendency of large values of components of a random vector to occur simultaneously. This kind of dependence information can be qualitatively different than what is given by numerical summaries, such as correlation, which average over the total body of the joint distribution. Two examples of the type of questions that extremal dependence analysis deals with are

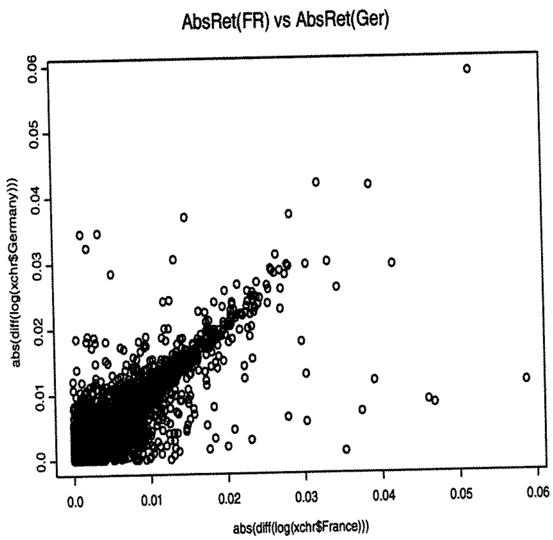
- Is a large movement in exchange rate returns in one currency, such as the German mark relative to the US dollar, likely to be accompanied by a similar large movement in another currency, such as the French frank? (See [43, 44].)
- In internet traffic, are large file sizes likely be imply a large transmission duration? (See [26, 5].)

Asymptotic independence for a bivariate vector means the probability of both components being large is of smaller order than the probability of one of them being large. If the components of the vector are non-negative, bivariate data from the distribution of such a vector has a scatter plot with data tending to hug the axes because if both components are unlikely to be simultaneously big, there will be few data points in the interior of the positive quadrant.

To illustrate the point, consider the the following example from [35]. The file fm-exch1.dat included with the program *Xtremes* (cf [31]), gives daily spot exchange rates of the currencies of France, Germany, Japan, Switzerland and the UK against the US dollar over a period of 6041 days from January 1971 to February 1994. Note the observation period is well before the introduction of the Euro.

Key words and phrases. heavy tails, multivariate regular variation, Pareto tails, asymptotic independence, extremal dependence.

Sidney Resnick's research was partially supported by NSF grant DMS-97-04982 at Cornell University.



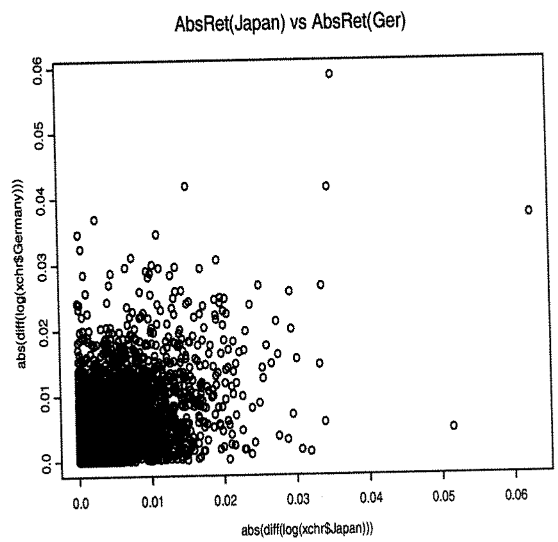


FIGURE 1. Scatter plots of absolute returns of (left) the French Franc against the German mark and (right) absolute returns of the Japanese Yen against the German mark.

Figure 1 gives on the left, a scatter plot of the absolute log daily returns for the French Franc against the absolute log daily returns for the German mark. Small log absolute returns for one currency are matched by a wide range of values for the other currency. Visually, however, dependence increases as the size of the absolute returns for the pair increases. The pattern varies, however, between different exchange rate processes. The dependence among large daily absolute returns between Japan and Germany (right) is much less pronounced than between France and Germany.

It is interesting to note that the correlation between the original exchange rate data for Germany and France is 0.579. The correlation between Japan and Germany is higher, namely 0.882. The large dependence between Germany and Japan as measured by correlation is not reflected in the scatterplot indicated in the right side of Figure 1 which indicates less dependence between extremes. The smaller correlation between France and Germany does not indicate the stronger dependence shown in the left plot of Figure 1 for the large absolute values of log returns.

In the internet traffic context, the dependence structure of large values of file size, transmission duration of the file and transmission throughput (file size divided by duration) has been analyzed in [26] and [5]. Correlations are emphasized in [46].

The data shown in Figure 2 are based on HTTP responses, gathered from the University of North Carolina main link during April of 2001 in a measurement study initiated by Kevin Jaffay and Don Smith (CS, UNC). An HTTP "response" is set of packets associated with a single HTTP data transfer, and "duration" is the time between the first and last packets. Packets were gathered over 21 four hour blocks, over each of the 7 days of the week, and

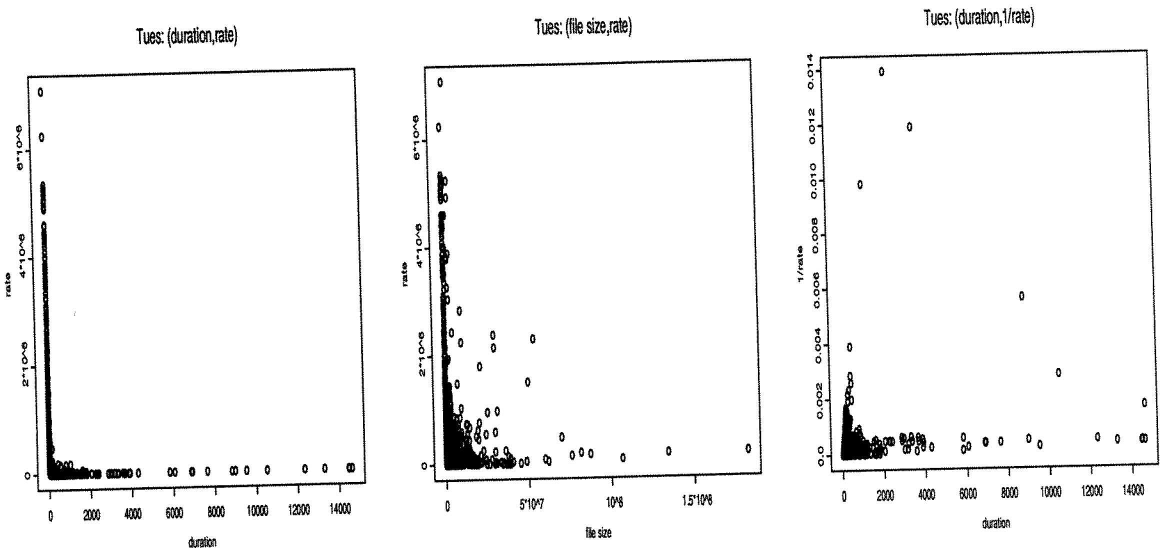


FIGURE 2. Scatter plots of (left) duration vs rate, (middle) file size vs rate and (right) duration vs inverse rate.

for "morning" (8:00AM-12:00AM), "afternoon" (1:00PM-5:00PM) and "evening" (7:30PM-11:30PM) periods on each day. The total number of HTTP flows over the four hour blocks ranged from $\sim$ 1 million (weekend mornings) to $\sim$ 7 million (weekday afternoons). Here we only consider Tuesday afternoon large flows, meaning thresholded data restricted to responses with more than 100 kilobytes.

The left plot in Figure 2 corresponding to duration vs throughput rate seems to exhibit clear axis hugging meaning there is little tendency for large durations to be associated with large rates. The middle plot for file size vs rate seems to exhibit some similar tendency. The right plot of duration vs inverse rate seems to exhibit extremal dependence. WARNING: Asymptotic independence is an asymptotic distributional property requiring scales for each variable to be adjusted appropriately and merely looking at scatter plots is not adequate. More careful study is required. In [26] and [5] it was found that the strongest tendency towards extremal independence was in the pair (file size, rate).

In Section 2 we review the notion of multivariate regular variation which underpins the theory of multivariate heavy tailed analysis. We review asymptotic independence and asymptotic full dependence and the statistical goal is to detect these situations from data. We also extend these ideas to stochastic processes and define a *regularly varying process*, which is an abuse of language since, of course, it is the distribution of the process with the regular variation property. Section 3 restricts attention to vectors of dimension 2 and defines the *extremal dependence measure* (EDM) which is standardized to have the look and feel of correlation. When the EDM is 0, we have asymptotic independence and when the EDM is 1, asymtotic full dependence is present. The EDM can also be used as a diagnostic for regularly varying processes and indicates independence properties between large values which are separated by sufficient time lags. The EDM was used in [5] as an exploratory data analysis tool and our goal is to begin the study of its mathematical properties.

Section 4 discusses estimators for parametric and semi-parametric quantities related to multivariate heavy tailed analysis and in particular suggests an estimator of the EDM. Asymtpotic normality for the EDM estimator is proven in Section 5.

In practice, heavy tailed vectors rarely, if ever, have component random variables with the same regular variation indices. None-the-less, this paper restricts attention to the *standard case*, which assumes each component random variable is tail equivalent to the others and each component variable has a distribution tail which is regularly varying with index -1. This assumption makes the probabilistic analysis and exposition clearer but begs the statistical question of how to transform a non-standard case to standard. There are two suggested methods to accomplish this involving either power transformations or ranks (see [5, 26, 33, 35, 34, 21, 23, 13, 42, 45, 30, 24, 25]) but the supremacy of either method in practice is not yet completely clear and proving asymptotic normality of the EDM without the standard case assumption is more difficult. So we have assumed the relatively easy case and more work remains to be done. We hope to address this soon.

2. Standard case multivariate regular variation.

Before getting to the meat and potatoes, here is a list of notational conventions that will make exposition easier.

2.1. Vector notation review. Vectors are denoted by bold letters, capitals for random vectors and lower case for non-random vectors. For example:

$$\boldsymbol{x} = (x^{(1)}, \dots, x^{(d)}) \in \mathbb{R}^d.$$

Operations between vectors *should (almost) always be interpreted componentwise* so that for two vectors $\boldsymbol{x}$ and $\boldsymbol{z}$

$$\begin{aligned}
\boldsymbol{x} < \boldsymbol{z} &\text{ means } x^{(i)} < z^{(i)},\ i = 1, \dots, d,\\
\boldsymbol{x} \leq \boldsymbol{z} &\text{ means } x^{(i)} \leq z^{(i)},\ i = 1, \dots, d,\\
\boldsymbol{x} = \boldsymbol{z} &\text{ means } x^{(i)} = z^{(i)},\ i = 1, \dots, d,\\
\boldsymbol{z}\boldsymbol{x} &=(z^{1)}x^{(1)}, \dots, z^{(d)}x^{(d)},\\
\boldsymbol{x} \bigvee \boldsymbol{z} &=(x^{(1)} \vee z^{(1)}, \dots, x^{(d)} \vee z^{(d)},\\
\frac{\boldsymbol{x}}{\boldsymbol{z}} &=\Big(\frac{x^{(1)}}{z^{(1)}}, \dots, \frac{x^{(d)}}{z^{(d)}}\Big)\\
\boldsymbol{x}^{\boldsymbol{z}} &=\big((x^{(1)})^{z^{(1)}}, \dots, (x^{(d)})^{z^{(d)}}\big).
\end{aligned}$$

Also $\mathbf{0} = (0, \dots, 0)$, $\mathbf{1} = (1, \dots, 1)$, $\boldsymbol{e}_i = (0, \dots, 1, \dots, 0)$ for $i = 1, \dots, d$. For a real number c, denote as usual

$$c\boldsymbol{x} = (cx^{(1)}, \dots, cx^{(d)}).$$

We denote rectangles by

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{R} : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$$

so that for $\boldsymbol{x} > \boldsymbol{0}$ and $\mathbb{E} = [\boldsymbol{0}, \boldsymbol{\infty}] \setminus \{\boldsymbol{0}\}$,

$$[\boldsymbol{0}, \boldsymbol{x}]^c = \mathbb{E} \setminus [\boldsymbol{0}, \boldsymbol{x}] = \{\boldsymbol{y} \in \mathbb{E} : \bigvee_{i=1}^{d} \frac{y^{(i)}}{x^{(i)}} > 1\}.$$

2.2. **Multivariate regularly varying functions.** A subset $C \subset \mathbb{R}^d$ is a *cone* if whenever $\boldsymbol{x} \in C$ also $t\boldsymbol{x} \in C$ for any $t > 0$. A function $h : C \mapsto (0, \infty)$ is monotone if it is either non-decreasing in each component or non-increasing in each component. For h non-decreasing, this is equivalent to saying that whenever $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\boldsymbol{x} \leq \boldsymbol{y}$ we have $h(\boldsymbol{x}) \leq h(\boldsymbol{y})$. The natural domain for a multivariate regularly varying function is a cone.

Suppose $h(\cdot) \geq 0$ is measurable and defined on C. Suppose $\boldsymbol{1} = (1, \dots, 1) \in C$. Call h *multivariate regularly varying on C with limit function* $\lambda(\cdot)$ if $\lambda(\boldsymbol{x}) > 0$ for $\boldsymbol{x} \in C$ and for all $\boldsymbol{x} \in C$ we have

$$\lim_{t \to \infty} \frac{h(t\boldsymbol{x})}{h(t\boldsymbol{1})} = \lambda(\boldsymbol{x}). \tag{2.1}$$

Note $\lambda(\boldsymbol{1}) = 1$. A simple scaling argument shows that $\lambda(\cdot)$ is homogeneous:

$$\lambda(s\boldsymbol{x}) = s^{\rho}\lambda(\boldsymbol{x}), \quad s > 0,\ \boldsymbol{x} \in C,\ \rho \in \mathbb{R}. \tag{2.2}$$

See [39, 15, 14, 16, 18, 1, 2, 27].

For multivariate distributions F concentrating on $[0, \infty)^d =: [\boldsymbol{0}, \boldsymbol{\infty})$, it is ambiguous what we mean by *distribution tail*. The usual interpretation has been to consider $1 - F(\boldsymbol{x})$ for $\boldsymbol{x} \geq \boldsymbol{0}$ but $\boldsymbol{x} \neq \boldsymbol{0}$ and so it is required that

$$\lim_{t \to \infty} \frac{1 - F(t\boldsymbol{x})}{1 - F(t\boldsymbol{1})} = \lambda(\boldsymbol{x}). \tag{2.3}$$

It is awkward to deal with distribution functions and more natural to deal with measures.

2.3. **Multivariate regularly varying tail probabilities.** There are various equivalences which define multivariate regularly varying tail probabilities. We restrict attention to the case of random vectors with non-negative components. Suppose $\{\boldsymbol{Z}_n, n \geq 1\}$ are iid random elements of $\mathbb{R}_+^d$ with common distribution $F(\cdot)$. Recall $\mathbb{E} = [\boldsymbol{0}, \boldsymbol{\infty}] \setminus \{\boldsymbol{0}\}$ and $[\boldsymbol{0}, \boldsymbol{x}]^c = \mathbb{E} \setminus [\boldsymbol{0}, \boldsymbol{x}]$. Set $M_+(\mathbb{E})$ to be the space of positive Radon measures on $\mathbb{E}$ and $M_p(\mathbb{E})$ is the space of Radon point measures on $\mathbb{E}$. Define the measure on $(0, \infty]$

$$\nu_\alpha(x, \infty] = x^{-\alpha}, \quad x > 0,\ \alpha > 0.$$

Vague convergence of measures is denoted by $\xrightarrow{v}$.

Fix a norm $\|\cdot\|$ and with respect to this norm define the unit sphere

$$\aleph = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}.$$

Set $\aleph_+ = \aleph \cap \mathbb{E}$. Define the polar coordinate transformation $T : \mathbb{R}^d \setminus \{\boldsymbol{0}\} \mapsto (0, \infty) \times \aleph$ by

$$T(\boldsymbol{x}) = \left(\|\boldsymbol{x}\|, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right) =: (r, \boldsymbol{a}),$$

and the inverse transformation $T^{\leftarrow} : (0, \infty) \times \aleph \mapsto \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ by

$$T^{\leftarrow}(r, \boldsymbol{a}) = r\boldsymbol{a}.$$

Think of $\boldsymbol{a} \in \aleph$ as defining a direction and r telling how far in direction $\boldsymbol{a}$ to proceed. Since we excluded $\mathbf{0}$ from the domain of T, both T and $T^{\leftarrow}$ are continuous bijections.

When $d = 2$, it is customary, but not obligatory, to write $T(\boldsymbol{x}) = (r\cos\theta, r\sin\theta)$, where $0 \leq \theta \leq 2\pi$, rather than the more consistent notation $T(\boldsymbol{x}) = \big(r, (\cos\theta, r\sin\theta)\big)$. For a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ we sometimes write $T(\boldsymbol{X}) = (R, \boldsymbol{\Theta})$.

To deal with multivariate regular variation of tail probabilities, we have to consider a punctured space with a one-point un-compactification such as $[\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$. Equivalences in terms of polar coordinates are then problematic since the polar coordinate transformation is not defined on the lines through $\boldsymbol{\infty}$, so some sort of restriction argument is necessary. A different treatment of the polar coordinate transformation is given in [1, 2, 28]. For versions of the following see [39, 35].

Theorem 1 (Multivariate regularly varying tail probabilities). *Suppose $\{\boldsymbol{Z}_m, m \geq 1\}$ are iid $\mathbb{R}^d_+$-valued random vectors with common distribution F. The following statements are equivalent. (In each, we understand the phrase Radon measure to mean a not identically zero Radon measure. Also, repeated use of the symbols ν, $b(\cdot)$, $\{b_n\}$ from statement to statement, does not require these objects to be exactly the same in different statements. See Remark 1 after Theorem 1.)*

(1) *There exists a Radon measure ν on $\mathbb{E}$ such that*

$$\lim_{t\to\infty} \frac{1 - F(t\boldsymbol{x})}{1 - F(t\mathbf{1})} = \lim_{t\to\infty} \frac{\mathbb{P}[\frac{\boldsymbol{Z}_1}{t} \in [\mathbf{0}, \boldsymbol{x}]^c]}{\mathbb{P}[\frac{\boldsymbol{Z}_1}{t} \in [\mathbf{0}, \mathbf{1}]^c]} = \nu\Big([\mathbf{0}, \boldsymbol{x}]^c\Big),$$

for all points $\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty}) \setminus \{\mathbf{0}\}$ which are continuity points of $\nu([\mathbf{0}, \cdot]^c)$.

(2) *There exists a function $b(t) \to \infty$ and a Radon measure ν on $\mathbb{E}$ such that in $M_+(\mathbb{E})$*

$$t\mathbb{P}[\frac{\boldsymbol{Z}_1}{b(t)} \in \cdot\,] \xrightarrow{v} \nu, \quad t \to \infty.$$

(3) *There exists a sequence $b_n \to \infty$ and a Radon measure ν on $\mathbb{E}$ such that in $M_+(\mathbb{E})$*

$$n\mathbb{P}[\frac{\boldsymbol{Z}_1}{b_n} \in \cdot\,] \xrightarrow{v} \nu, \quad t \to \infty.$$

(4) *There exists a probability measure $S(\cdot)$ on $\aleph_+$ and a function $b(t) \to \infty$ such that for $(R_1, \boldsymbol{\Theta}_1) = \Big(\|\boldsymbol{Z}_1\|, \frac{\boldsymbol{Z}_1}{\|\boldsymbol{Z}_1\|}\Big)$ we have*

$$t\mathbb{P}[(\frac{R_1}{b(t)}, \boldsymbol{\Theta}_1) \in \cdot\,] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+\big(((0, \infty] \times \aleph_+\big)$, where $c > 0$.

(5) *There exists a probability measure $S(\cdot)$ on $\aleph_+$ and a sequence $b_n \to \infty$ such that for $(R_1, \boldsymbol{\Theta}_1) = \Big(\|\boldsymbol{Z}_1\|, \frac{\boldsymbol{Z}_1}{\|\boldsymbol{Z}_1\|}\Big)$ we have*

$$n\mathbb{P}[(\frac{R_1}{b_n}, \boldsymbol{\Theta}_1) \in \cdot\,] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+\big((0, \infty] \times \aleph_+\big)$, where $c > 0$.

(6) *There exists $b_n \to \infty$ such that in $M_p(\mathbb{E})$*

$$\sum_{i=1}^{n} \epsilon_{\boldsymbol{Z}_i/b_n} \Rightarrow PRM(\nu),$$

where $PRM(\nu)$ is a Poisson random measure with mean measure ν.

(7) *There exists a sequence $b_n \to \infty$ such that in $M_p((0,\infty] \times \aleph_+)$*

$$\sum_{i=1}^{n} \epsilon_{(R_i/b_n, \Theta_i)} \Rightarrow PRM(c\nu_\alpha \times S).$$

These conditions imply that for any sequence $k = k(n) \to \infty$ such that $n/k \to \infty$ we have

(8) *In $M_+(\mathbb{E})$*

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\boldsymbol{Z}_i/b(\frac{n}{k})} \Rightarrow \nu \tag{2.4}$$

and 8 is equivalent to any of 1–7, provided $k(\cdot)$ satisfies $k(n) \sim k(n+1)$. Similar statements to (2.4) can be made in terms of polar coordinates.

Remark 1. Normalization of all components by the same function means that marginal distributions are tail equivalent; that is, [32, 39]

$$\lim_{x\to\infty} \frac{P[Z_1^{(i)} > x]}{P[Z_1^{(j)} > x]} =: r_{ij} \in [0,\infty],$$

for $1 \le i,j \le d$. To avoid cases where some marginal tails are heavier than others, corresponding to $d_{ij} = 0$ or ∞ for some (i,j), we usually assume all components $\{Z_1^{(i)}, 1 \le i \le d\}$ are identically distributed.

When $b(t) = t$ or $b_n = n$ we are in the *standard case* [17, 39] and all marginal distributions are tail equivalent to a standard Pareto distribution with $\alpha = 1$. In general, the possible choices of $b(t)$ include

(i) $b(t) = \left(\frac{1}{1-F_{(1)}}\right)^{\leftarrow}(t)$ where $F_{(1)}(x) = P[Z_1^{(1)} \le x]$ is the one-dimensional marginal distribution.

(ii) $b(t) = \left(\frac{1}{1-F_R}\right)^{\leftarrow}(t)$ where $F_R(x) = P[R_1 \le x]$ is the distribution of $\|\boldsymbol{Z}_1\|$. Note this choice of $b(\cdot)$ depends on the choice of norm $\|\cdot\|$.

Different choices of $b(\cdot)$ may introduce different constants c in the limit statements.

Definition 1. A non-negative stationary stochastic sequence $\{X_n, n \ge 0\}$ is called a regularly varying process if for every $d \ge 0$, there exists a Radon measure $\nu_{0,\dots,d}(\cdot)$ on $[0,\infty]^{d+1} \setminus \{\boldsymbol{0}\}$ such that

$$nP[b_n^{-1}(X_0,\dots,X_d) \in \cdot\,] \xrightarrow{v} \nu_{0,\dots,d}(\cdot), \tag{2.5}$$

where we assume b_n is chosen to satisfy

$$nP[X_1 > b_n x] \to x^{-\alpha}, \quad x > 0.$$

An analogous definition can be made in continuous time. Examples of continuous time regularly varying processes are stationary stable processes ([40, 41]) and stationary max-stable processes ([12, 20, 11]). In discrete time there are moving and max-moving averages with heavy tailed innovations as well as autoregressions and ARMA's. Stationary ARCH and GARCH processes are also regularly varying processes [28, 29, 2, 10].

2.4. **Asymptotic independence.** Suppose $\{\boldsymbol{Z}_n, n \geq 1\}$ are iid and satisfy the conditions of Theorem 1. The distribution F of $\boldsymbol{Z}_1$ possesses the property of *asymptotic independence* if

(1) $\nu(\mathbb{E}^0) = 0$ so that ν concentrates on the axes where
$$\mathbb{E}^0 = \{bs \in \mathbb{E} : \text{ For some } 1 \leq i < j \leq d, s^{(i)} \wedge s^{(j)} > 0\};$$
OR

(2) S concentrates on $\{\boldsymbol{e}_i, i = 1, \ldots, d\}$, where recall
$$\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$$
is the i-th basis vector.

Note that if $d = 2$, then ν concentrates on the horizontal and vertical axes and S puts all mass on the two basis vectors on the two axes.

The distribution F of $\boldsymbol{Z}_1$ possesses the property of *asymptotic full dependence* if

(1) ν concentrates on $\{t\frac{\boldsymbol{1}}{\|\boldsymbol{1}\|} : t > 0\}$, the diagnonal line,
or

(2) S concentrates on $\{\frac{\boldsymbol{1}}{\|\boldsymbol{1}\|}\}$.

When $d = 2$ and components of the random vector are tail equivalent, asymptotic independence means
$$\mathbb{P}[Z_1^{(1)} > b_n | Z_1^{(2)} > b_n] = \frac{\mathbb{P}[Z_1^{(1)} > b_n, Z_1^{(2)} > b_n]}{\mathbb{P}[Z_1^{(2)} > b_n]} \to \nu((\boldsymbol{1}, \boldsymbol{\infty})) = 0.$$
Hence the name, asymptotic independence. The extreme value background is discussed [39, page 290].

The goal is to detect asymptotic independence statistically.

3. The extremal dependence measure.

The extremal dependence measure is a crude measure of dependence between large values of the various components of a random vector. It is convenient to suppose $d = 2$; higher dimensional analogues are possible but it is dubious how statistically useful such analogues would be.

Suppose $\boldsymbol{Z} = (Z^{(1)}, Z^{(2)})$ is a bivariate heavy tailed vector whose distribution F satisfies the conditions of Theorem 1. For $d = 2$, assume the angular measure S is defined on $[0, \pi/2]$. Define
$$v := \int_0^{\pi/2} (\theta - \frac{\pi}{4})^2 S(d\theta). \tag{3.1}$$

The distribution of $\boldsymbol{Z}$ possesses asymptotic full dependence iff $v = 0$ and asymptotic independence iff $v = (\pi/4)^2$. These facts follow easily since the integral in (3.1) is extreme when S concentrates all mass on the atoms $\{0, \pi/2\}$. The extremal dependence measure is defined by

$$\rho := 1 - \frac{v}{(\pi/4)^2}, \tag{3.2}$$

and the distribution of $\boldsymbol{Z}$ possesses asymptotic independence iff $\rho = 0$ and asymptotic full dependence iff $\rho = 1$.

Suppose $\{X_n, n \geq 0\}$ is a regularly varying process defined in (2.5). We define

$$\rho(l) = \text{ the EDM of } (X_0, X_l), \tag{3.3}$$

which we write in shorthand as

$$\rho(l) = \mathrm{EDM}(X_0, X_l).$$

Return to the case $d = 2$ and assume Theorem 1 holds. Then

$$\bigvee_{i=1}^{n} \frac{\boldsymbol{Z}_i}{b_n} \Rightarrow \left(M^{(\infty)}(1), M^{(\infty)}(2)\right)$$

where the limit is a max-stable random vector. If the EDM is 0, the components of the limit are independent. The next Theorem helps explain how the EDM controls simultaneous large values in a stationary regularly varying sequence. (Cf. [4, Proposition 3.2.1, page 89].)

Remark 2. The limit measure ν is associated with *spectral functions* $f_1, f_2 \in L_1([0,1], ds)$ which are integrable on $[0,1]$ with respect to Lebesgue measure. See, for example, [39, Proposition 5.11, page 268]. In terms of these functions, the integral v in (3.1) can be expressed as

$$v = \int_0^1 \left| \arctan \frac{f_2(s)}{f_1(s)} - \arctan 1 \right|^2 ds.$$

Theorem 2. *Suppose $\boldsymbol{X} := \{X^{(n)}, n \geq 0\}$ is a regularly varying sequence satisfying (2.5) whose EDM has the property that*

$$\rho(l) = EDM(X^{(0)}, X^{(l)}) = 0, \text{ for } l \geq L. \tag{3.4}$$

Let $\{\boldsymbol{X}(j), j \geq 1\}$ be iid $\mathbb{R}_+^\infty$ valued random elements with $\boldsymbol{X}(j) \stackrel{d}{=} \boldsymbol{X}$. Then in $\mathbb{R}_+^\infty$,

$$\frac{\bigvee_{j=1}^{n} \boldsymbol{X}(j)}{b_n} \Rightarrow \boldsymbol{M}(\infty), \tag{3.5}$$

where $\boldsymbol{M}(\infty) = (M^{(1)}(\infty), M^{(2)}(\infty), \dots)$ is a max-stable process with the property that if I_1, I_2 are two finite subsets of the non-negative integers, $I_m \subset \mathbb{N} = \{0, 1, \dots\}$, $m = 1, 2$ and $I_1 \cap I_2 = \emptyset$ and $\inf I_2 - \sup I_1 \geq L$ then

$$\{M^{(l)}(\infty), l \in I_1\} \text{ and } \{M^{(l)}(\infty), l \in I_2\}$$

are independent.

Proof. The limit sequence must be a stationary max-stable sequence. There exist (see, for example [39, page 268]) non-negative functions $f_j, g_m,\ j \in I_1, m \in I_2$ in $L_1([0,1], ds)$ such that (assuming $\alpha = 1$)

$$
\begin{aligned}
P[M^{(j)}(\infty) \leq x_j, j \in I_1;\ M^{(m)}(\infty) \leq x_m, m \in I_2]& \\
= \exp\{-\int_0^1 \Big(\bigvee_{j\in I_1} \frac{f_j(s)}{x_j} \bigvee \bigvee_{m \in I_2} \frac{g_m(s)}{x_m}\Big) ds\}.& \qquad (3.6)
\end{aligned}
$$

Note that for $i \in I_1,\ m \in I_2$

$$
f_i(s) g_m(s) = 0, \tag{3.7}
$$

almost everywhere since $\rho(l) = 0$ for $l \geq L$ and therefore $M^{(\infty)}(i)$ and $M^{(\infty)}(m)$ are independent.

Now write the exponent in (3.6) as

$$
\begin{aligned}
&\int_0^1 \Big(\bigvee_{j\in I_1} \frac{f_j(s)}{x_j} \bigvee \bigvee_{m \in I_2} \frac{g_m(s)}{x_m}\Big) ds \\
&\qquad = \int_{[\vee_{j\in I_1} f_j > 0]} + \int_{[\vee_{j \in I_1} f_j = 0]}
\end{aligned}
$$

and because of (3.7), this is the same as

$$
\begin{aligned}
&= \int_{[\vee_{j\in I_1} f_j > 0]} \bigvee_{j \in I_1} \frac{f_j(s)}{x_i} ds + \int_{[\vee_{j\in I_1} f_j = 0]} \bigvee_{m \in I_2} \frac{g_m(s)}{x_m} ds \\
&= \int_{[\vee_{j\in I_1} f_j > 0]} \bigvee_{j \in I_1} \frac{f_j(s)}{x_i} ds + \int_{[\vee_{m\in I_2} g_m > 0]} \bigvee_{m \in I_2} \frac{g_m(s)}{x_m} ds
\end{aligned}
$$

which is of the form

$$
h_1(x_j, j \in I_1) + h_2(x_m, m \in I_2). \tag{3.8}
$$

This suffices to show the independence. $\square$

Corollary 1. *Assume the notation and conditions of Theorem 2. Let $|I_j|$ be the number of elements in I_j. Then in $M_p\big([0,\infty]^{|I_1|+|I_2|} \setminus \{\mathbf{0}\}\big)$ we have*

$$
\sum_{i=1}^n \epsilon_{b_n^{-1}\left(X^{(j)}(i), j\in I_1; X^{(m)}(i), m \in I_2\right)} \Rightarrow N_1 + N_2 \tag{3.9}
$$

where N_1 and N_2 are independent PRM's of the form

$$
\begin{aligned}
N_1 &= \sum_k \epsilon_{\left(j^{(l)}(k),\ l \in I_1, \mathbf{0}\right)}, \\
N_2 &= \sum_k \epsilon_{\left(\mathbf{0},\ j^{(m)}(k),\ m \in I_2\right)}.
\end{aligned}
$$

Proof. The conclusion (3.5) of Theorem 2 implies

$$nP\big[b_n^{-1}\big(X^{(j)}(i), j \in I_1; X^{(m)}(i), m \in I_2\big) \in [\mathbf{0}, \boldsymbol{x}]^c\big] \to h_1(x_j, j \in I_1) + h_2(x_m, m \in I_2)$$

using the notation of (3.8). The stated result now follows from [39, Proposition 3.21, page 154]. □

Corollary 2. *Assume the notation and conditions of Theorem 2. Suppose* $0 < \alpha < 2$. *Then there exists* c_n *such that in* $\mathbb{R}_+^\infty$

$$\sum_{i=1}^{n} \frac{\boldsymbol{X}(i)}{b_n} - nc_n \Rightarrow \boldsymbol{X}_\alpha,$$

where $\boldsymbol{X}_\alpha$ *is a stationary* α*-stable sequence such that*

$$\{X_\alpha^{(i)}, i \in I_1\} \text{ and } \{X_\alpha^{(m)}, m \in I_2\}$$

are independent.

Proof. Convergence of finite dimensional distributions is proven as in [36]. The sequence $\{X_\alpha^{(i)}, i \in I_p\}$ is constructed from N_p, $p = 1, 2$. Independence follows from Corollary 1. □

4. Estimators and diagnostics.

Suppose the conditions of Theorem 1 hold. The analogue of (2.4) provides a way to estimate the spectral measure S:

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{R_i}{b(n/k)}, \Theta\right)} \Rightarrow c\nu_\alpha \times S. \tag{4.1}$$

as $k = k(n) \to \infty$ and $k/n \to 0$. There are (at least) three problems:

- What is c? This can be scaled away either by taking ratios in (4.1) or by assuming $b(t) = \left(\frac{1}{1-F_R}\right)^{\leftarrow}(t)$, since then $tP[R_1 > b(t)] \to 1$ as $t \to \infty$.
- How does one estimate $b(\cdot)$? In the standard case $b(t) = t$ but if Theorem 1 is assumed to hold with general $b(\cdot)$, we can replace $b(n/k)$ with $\hat{b}(n/k)$, the k-th largest order statistic of the iid sample $R_1, \dots, R_n$. (In practice, if we have transformed a non-standard case to standard by power transformation, the order statistic normalization seems to work better than normalizing by n/k.)
- How does one choose k? There is a graphical technique due to Stărică [43, 34, 35] which seems to work reasonably well which requires examining scaling plots and some trial and error. This technique is exploratory and to-date, nothing is proven about it.

Specialize to $d = 2$. Let $\{\boldsymbol{Z}_j, 1 \le j \le n\}$ be iid random pairs. Using an estimator $\hat{b}(\frac{n}{k})$, define

$$\hat{\nu}(\cdot) = \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{\boldsymbol{Z}_i}{\hat{b}(n/k)}} \tag{4.2}$$

$$
\begin{aligned}
\hat{S}(\cdot) =& \frac{\hat{\nu}\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1, \arctan \frac{x^{(2)}}{x^{(1)}} \in \cdot\}}{\hat{\nu}\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1\}} \\
=& \frac{\sum_{i=1}^{n} 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k)>1]} \epsilon_{\Theta_i}(\cdot)}{\sum_{i=1}^{n} 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k)>1]}}
\end{aligned} \tag{4.3}
$$

where $\Theta_i = \arctan\big(Z_i^{(2)}/Z_i^{(1)}\big)$. Also define

$$
\begin{aligned}
\hat{v} =& \int_0^{\pi/2} (\theta - \pi/4)^2 \hat{S}(d\theta) \\
=& \frac{\sum_{i=1}^{n} 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k)>1]} (\Theta_i - \pi/4)^2}{\sum_{i=1}^{n} 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k)>1]}}
\end{aligned} \tag{4.4}
$$

and

$$
\hat{\rho} = 1 - \frac{\hat{v}}{(\pi/4)^2}. \tag{4.5}
$$

Provided in (4.2) that

$$
\hat{\nu} \Rightarrow \nu
$$

in $M_+([0,\infty]^2 \setminus \{\boldsymbol{0}\}$, as will certainly be the case if, as assumed, the $\boldsymbol{Z}$'s are iid, we get by continuous mapping that all other quantities are also consistent:

$$
\hat{S} \Rightarrow S, \quad \hat{v} \xrightarrow{P} v, \quad \hat{\rho} \xrightarrow{P} \rho.
$$

5. Asymptotic normality of the EDM estimator.

Continue to suppose $d = 2$ and that Theorem 1 holds. Assume $\alpha = 1$ and that either $b(\cdot)$ is known, perhaps because the the standard case assumption holds. It is most convenient to assume

$$
tP[R_1 > b(t)] \to 1, \quad (t \to \infty). \tag{5.1}
$$

Since we assume we know $b(\cdot)$, there is no need for an estimate and we consider the quantities in (4.2)–(4.5) with $b(\cdot)$ instead of $\hat{b}(\cdot)$. Define

$$
N_n = \sum_{i=1}^{n} 1_{[R_i > b(n/k)]} \tag{5.2}
$$

as the random number of exceedances. Let $\{i(l,n), l \geq 1\}$ be the random indices such that $R_{i(l,n)} > b(n/k)$ so that ([39, page 212]) $\{\Theta_{i(l,n)}, l \geq 1\}$ are iid, independent of $\{N_n\}$ and

$$
\begin{aligned}
P[\Theta_{i(1,n)} \in \cdot] =& P[\Theta_1 \in \cdot | R_1 > b(n/k)] \\
=& \frac{P[R_1 > b(n/k), \Theta_1 \in \cdot]}{P[R_1 > b(n/k)]} \\
\to& S(\cdot),
\end{aligned}
$$

so that for an iid sequence $\{\Theta_l^{(\infty)}, l \geq 1\}$ with common distribution S we have

$$\{\Theta_{i(l,n)}, l \geq 1\} \Rightarrow \{\Theta_l^{(\infty)}, l \geq 1\}, \quad n \to \infty.$$

This allows us to represent $\hat{v}$ as a random sum of iid random variables. From (4.4)

$$\hat{v} = \frac{1}{N_n} \sum_{l=1}^{N_n} (\Theta_{i(l,n)} - \pi/4)^2. \tag{5.3}$$

Now we assume asymptotic independence. From asymptotic independence we observe that, as $n \to \infty$,

$$E(\Theta_{i(1,n)} - \pi/4)^2 \to \int_0^{\pi/2} (\theta - \pi/4)^2 S(d\theta) = \left(\frac{\pi}{4}\right)^2,$$

and

$$\mathrm{Var}\Big((\Theta_{i(1,n)} - \pi/4)^2\Big) \to 0, \tag{5.4}$$

the last line following from the fact that $(\Theta_1^{(\infty)} - \pi/4)^2$ is almost surely constant with respect to the two-point distribution on $\{0, \pi/2\}$, the constant being $(\pi/2 - \pi/4)^2 = (\pi/4)^2$ or $(0 - \pi/4)^2 = (\pi/4)^2$.

Theorem 3. *Suppose $\{\boldsymbol{Z}_n, n \geq 1\}$ is iid with common distribution F and that $d = 2$ and Theorem 1 holds as well as* (5.1)*. Assume also that asymptotic independence holds so that S is a two-point distribution concentrating mass on $\{0, \pi/2\}$. Finally, suppose*

$$Var\Big(((\Theta_1 - \pi/4)^2) | R_1 > b(n/k)\Big) \neq 0, \tag{5.5}$$

for $n \geq n_0$ for some n_0. Then

$$\frac{\sqrt{k}}{\sqrt{Var\Big((\Theta_{i(1,n)} - \pi/4)^2\Big)}} \Big(\hat{v} - E\big((\Theta_{i(1,n)} - \pi/4)^2\big)\Big) \Rightarrow W(1), \quad n \to \infty, \tag{5.6}$$

where $\{W(t), t \geq 0\}$ is a standard Wiener process. Consequently,

$$\frac{\sqrt{k}}{\sqrt{Var\Big((\Theta_{i(1,n)} - \pi/4)^2\Big)}} \left(\hat{\rho} - \left[1 - \frac{E(\Theta_{i(1,n)} - \pi/4)^2}{(\pi/4)^2}\right]\right) \Rightarrow \frac{-W(1)}{(\pi/4)^2} \stackrel{d}{=} \frac{W(1)}{(\pi/4)^2} \tag{5.7}$$

REMARKS.

(1) It is important to note that in the asymptotic independence case, the rate of convergence is not k but

$$\frac{\sqrt{k}}{\sqrt{\mathrm{Var}\Big((\Theta_{i(1,n)} - \pi/4)^2\Big)}}$$

where the denominator is converging to zero. Without asymptotic independence, the denominator would converge to a constant and the rate of convergence would be $\sqrt{k}$.

This is the essential difference between the asymptotic independence case and cases without the asymptotic independence.

(2) If (5.5) fails then there exists a sequence $n_p, p \geq 1$ such that
$$P[|\Theta_{i(1,n_p)} - \frac{\pi}{4}| = \sqrt{c_{n_p}}] = 1,$$
where $c_n \to (\pi/4)^2$. This could not happen if, for instance, the underlying distribution F had a density.

Proof. From (5.2) and the Law of Large Numbers or the analogue of (2.4)
$$\frac{N_n}{k} \Rightarrow 1, \tag{5.8}$$
in $[0,\infty)$.

From the functional central limit theorem for triangular arrays
$$\frac{1}{\sqrt{k}\sqrt{\operatorname{Var}\Big((\Theta_{i(1,n)} - \pi/4)^2\Big)}} \sum_{l=1}^{[kt]} \Big((\Theta_{i(l,n)} - \pi/4)^2 - E\big((\Theta_{i(l,n)} - \pi/4)^2\big)\Big) \Rightarrow W(t), \tag{5.9}$$
in $D[0,\infty)$. Because $\{\Theta_{i(l,n)}, l \geq 1\}$ are independent of $\{N_n\}$, the statements (5.8) and (5.9) can be combined into a joint convergence statement. Apply the almost surely continuous map $(f(\cdot), c) \mapsto f(c)$ from $D[0,\infty) \times [0,\infty) \mapsto [0,\infty)$ and we get
$$\frac{1}{\sqrt{k}\sqrt{\operatorname{Var}\Big((\Theta_{i(1,n)} - \pi/4)^2\Big)}} \sum_{l=1}^{N_n} \Big((\Theta_{i(l,n)} - \pi/4)^2 - E\big((\Theta_{i(l,n)} - \pi/4)^2\big)\Big) \Rightarrow W(1),$$
and since
$$N_n \hat{v} = \sum_{l=1}^{N_n} (\Theta_{i(l,n)} - \pi/4)^2$$
we conclude
$$\frac{N_n}{\sqrt{k}\sqrt{\operatorname{Var}\Big((\Theta_{i(1,n)} - \pi/4)^2\Big)}} \Big(\hat{v} - E\big((\Theta_{i(l,n)} - \pi/4)^2\big)\Big) \Rightarrow W(1).$$
Since $N_n/k \xrightarrow{P} 1$, the result follows. □

Of course, we would prefer the centering in (5.6) to be $v = (\pi/4)^2$ and the centering in (5.5) to be zero but then we would have one sided random variables converging to the two-sided normal limit. Any change of centering would require a second order regular variation condition which would be difficult to check in practice. (See [19] for background discussion.) Attempts to derive a second order condition in polar coordinates from a second order condition in Cartesian coordinates are discussed in [33]. We will not pursue this here.

Of course, for statistical usage, we need, at least, an approximate solution to equations of the form
$$0.05 = P[\hat{\rho} > x_{0.05}].$$

Using (5.5) we have with the notation $P[W(1) > N^{\leftarrow}(0.95)] = 0.05$ that

$$x_{0.05} \approx \frac{N^{\leftarrow}(0.95)}{(\pi/4)^2 \frac{\sqrt{k}}{\sqrt{\mathrm{Var}\left((\Theta_{i(1,n)}-\pi/4)^2\right)}}} + 1 - \frac{E\left((\Theta_{i(1,n)} - \pi/4)^2\right)}{(\pi/4)^2}. \tag{5.10}$$

One would like to be able to replace $E\left((\Theta_{i(1,n)} - \pi/4)^2\right)$ by the plug-in-estimator consisting of the sample average $\frac{1}{N_n}\sum_{l=1}^{N_n}(\Theta_{i(l,n)} - \pi/4)^2$ and this requires analysis of the difference between the two terms. Similarly, one would like to be able to replace $\mathrm{Var}\left((\Theta_{i(1,n)} - \pi/4)^2\right)$ by the plug-in-estimator consisting of $\widehat{\mathrm{Var}}$, the sample variance of $\{(\Theta_{i(l,n)} - \pi/4)^2, 1 \leq l \leq N_n\}$ and justifying this requires

$$\frac{\widehat{\mathrm{Var}}}{\mathrm{Var}\left((\Theta_{i(1,n)} - \pi/4)^2\right)} \xrightarrow{P} 1,$$

as $n \to \infty$.

6. Examples and concluding remarks

Theorems 2 and 3 indicate that one ought to be able to treat a regularly varying sequence as a time series and test for asymtptotic independence beyond some lag. We have experimented with doing this and the results promising but require further effort and justification. In general we do not know $b(\cdot)$ and so we replace $b(\cdot)$ by the k-th largest order statistic of the norms. This requires choice of k and experimentation with various values. In (5.10) the mean and variance are replaced by the sample versions and this needs justification. Finally, in the time series context, if we suspect asymptotic independence beyone lag L of the time series $\{X_j, 1 \leq j \leq n\}$, then we analyze pairs $\{(X_j, X_{j+L}), 1 \leq j \leq n - L\}$ and it is unlikely that for different values of j all pairs are independent as assumed in Theorem 3.

Undaunted by these difficulties, here are the results of some experiments.

1. Moving averages. We constructed a moving average of order 6 using equal weights applied to 100,000 standard Pareto random variables with $\alpha = 1$. This is a case where beyond lag 6, variables are independent as well as asymptotically independent.

On the left of Figure 3 is the EDM plot $(l, EDM(X^{(0)}, X^{(l)}, 1 \leq l \leq 25)$ of MA as a function of lag using $k = 1000$. The plot certainly captures the asymptotic independence beyond lag 6. The horizontal line is given by (5.10). For comparison, the sample correlation plot of MA is given on the right of Figure 3. Moving average processes are one of the few classes of regularly varying processes for which the acf plot is informative ([37]).

2. Mixtures. We next constructed a regularly varying sequence called *notindep* that was dependent but possessed asymptotic independence. We did this by taking 20,000 iid standard Pareto observations with $\alpha = 1$ and multiplying the whole collection by a single independent Pareto observation with $\alpha = 2$. By a generalization ([38, 28]) of Breiman's theorem [3], this produces a regularly varying sequence whose marginal distribution has $\alpha = 1$ and it is easy to check that the sequence is asymptotically independent.

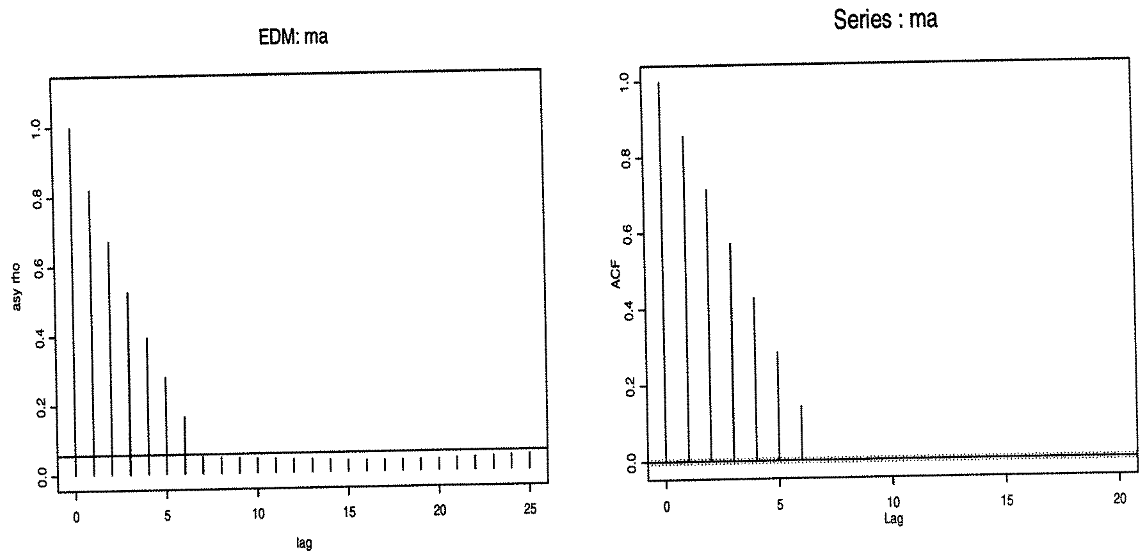


FIGURE 3. EDM plot of MA as a function of the lag (left) and acf plot of MA (right).

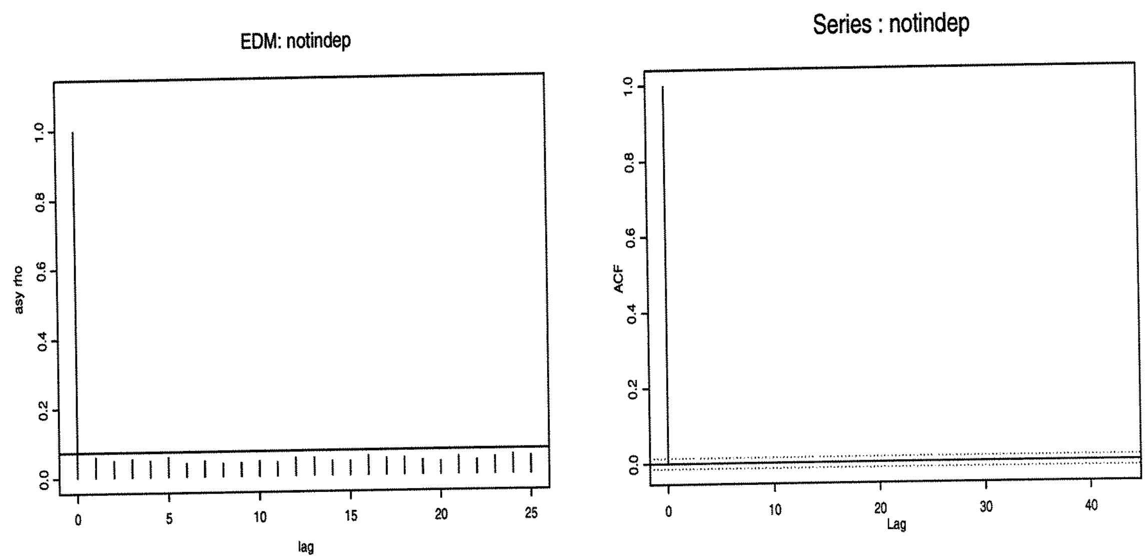


FIGURE 4. EDM plot of *notindep* as a function of the lag (left) and acf plot of *notindep* (right).

Figure 4 displays the EDM plot on the left for *notindep*. The small values of $\{\rho(l), 1 \leq l \leq 20\}$ show strong tendency towards asymptotic independence. The ACF-plot on the right is expected but note that this plot fails to capture the lack of independence since the ACF thinks the data is uncorrelated (even though correlations do not exist).

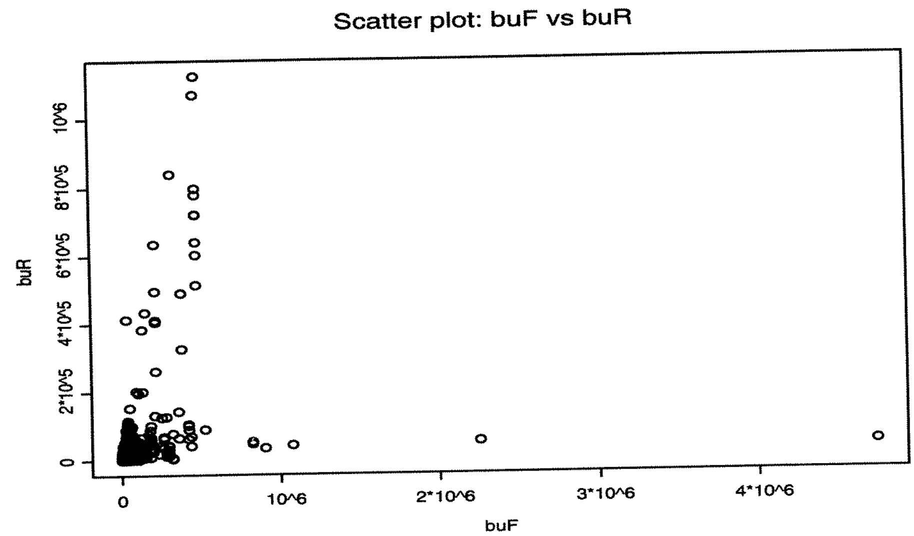


FIGURE 5. Scatter plot of buF vs buR.

3. BU DATA. The Boston University data is extensively documented and studied (see [9, 6, 7, 8, 22, 26]) and is available at `www.acm.org/sigcomm/ITA/`, the Internet Traffic Archive (ITA) web site. This data used here is processed from the original 1995 Boston University data and consists of 4161 file sizes (F) and transmission rates (R) inferred from the time required for downloading a file and the file size. The data, thus, consists of bivariate pairs (F,R) and a scatter plot is shown in Figure 5.

Each marginal distribution is heavy tailed and a combination of Hill and QQ plotting estimates the alphas as (1.157,1.138) for F and R respectively. Then we raise each component to its α-power to transform to the standard case where each marginal is regularly varying with unit α. We then compute $\hat{\rho}$ and the quantile $x_{0.05}$ in (5.10) and find with $k = 1000$ that

$$\hat{\rho} = 0.579 \text{ and } x_{0.05} = 0.614$$

and thus, despite the relatively large value of $\hat{\rho}$, this analysis presents no evidence at the 0.05 level against the hypothesis of asymptotic independence. Experimenting with a range of k-values produces the same conclusion.

7. ACKNOWLEDGEMENT

Krishanu Maulik made many helpful comments on an early draft of this paper which corrected some sloppy errors and led to several improvements.

REFERENCES

[1] B. Basrak. *The sample autocorrelation function of non-linear time series.* PhD thesis, Rijksuniversiteit Groningen, Groningen, Netherlands, 2000.

[2] B. Basrak, R. Davis, and T. Mikosch. A characterization of multivariate regular variation. Preprint, 2000.

[3] L. Breiman. On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.*, 10:323–331, 1965.

[4] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, second edition, 1991.

[5] F. H. Campos, J.S. Marron, S. Resnick, and K. Jaffay. Extremal dependence: Internet traffic applications. Web available at http://www.orie.cornell.edu/~sid, 2002.

[6] M. Crovella and A. Bestavros. Explaining world wide web traffic self–similarity. Preprint available as TR-95-015 from {crovella,best}@cs.bu.edu, 1995.

[7] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. In *Proceedings of the ACM SIGMETRICS '96 International Conference on Measurement and Modeling of Computer Systems*, volume 24, pages 160–169, 1996.

[8] M. Crovella, A. Bestavros, and M.S. Taqqu. Heavy-tailed probability distributions in the world wide web. In M.S. Taqqu R. Adler, R. Feldman, editor, *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions.* Birkhäuser, Boston, 1999.

[9] C. Cunha, A. Bestavros, and M. Crovella. Characteristics of www client–based traces. Preprint available as BU-CS-95-010 from {crovella,best}@cs.bu.edu, 1995.

[10] R.A. Davis and T. Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statistics.*, 26(5):2049–2080, 1998.

[11] Richard A. Davis and Sidney I. Resnick. Prediction of stationary max-stable processes. *Ann. Applied Probability*, 3(2):497–525, 1993.

[12] L. de Haan. A spectral representation for max-stable processes. *Ann. Probab.*, 12:1194–1204, 1984.

[13] L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–46, 1998.

[14] L. de Haan and E. Omey. Integrals and derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions II. *Stochastic Process. Appl.*, 16(2):157–170, 1984.

[15] L. de Haan, E. Omey, and S. Resnick. Domains of attraction and regular variation in $\mathbb{R}^d$. *J. Multivariate Anal.*, 14(1):17–33, 1984.

[16] L. de Haan and S. I. Resnick. Derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. *Stochastic Process. Appl.*, 8(3):349–355, 1978/79.

[17] L. de Haan and S.I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie*, 40:317–337, 1977.

[18] L. de Haan and S.I. Resnick. On regular variation of probability densities. *Stochastic Processes and their Applications*, 25:83–95, 1987.

[19] L. de Haan and S.I. Resnick. On asymptotic normality of the Hill estimator. *Stochastic Models*, 14:849–867, 1998.

[20] L.F.M. de Haan and James III Pickands. Stationary min-stable stochastic processes. *Probab. Theory Relat. Fields*, 72:477–492, 1986.

[21] J. Einmahl, L. de Haan, and V. Piterbarg. Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statistics*, 29(5):1401–1423, 2001.

[22] C.A. Guerin, H. Nyberg, O. Perrin, S. Resnick, H. Rootzen, and C. Stărică. Empirical testing of the infinite source poisson data traffic model. Technical Report 1257, School of ORIE, Cornell University, Ithaca NY 14853; available at www.orie.cornell.edu/~sid; to appear: Stochastic Models, 1999.

[23] Xin Huang. *Statistics of Bivariate Extreme values.* Phd thesis, Tinbergen Institute Research Series 22, Erasmus University Rotterdam, Postbus 1735, 3000DR, Rotterdam, The Netherlands, 1992.

[24] Anthony W. Ledford and Jonathan A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

[25] Anthony W. Ledford and Jonathan A. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, 59(2):475–499, 1997.

[26] K. Maulik, S. Resnick, and H. Rootzén . A network traffic model with random transmission rate. Technical report, web available at www.orie.cornell.edu/~sid; To appear: J. Applied Probability, 2003.

[27] Mark M. Meerschaert and Hans-Peter Scheffler. *Limit distributions for sums of independent random vectors.* John Wiley & Sons Inc., New York, 2001. Heavy tails in theory and practice.

[28] T. Mikosch. Modeling dependence and tails of financial time series. In *SemStat: Seminaire Europeen de Statistique, Exteme Value Theory and Applications*. Chapman-Hall, London, 2001. 75 pages; to appear.

[29] T. Mikosch and C. Stărică. Long range dependence effects and ARCH modeling. Available at: http://www.math.ku.dk/ mikosch/preprint.html, 2000.

[30] Ser-Huang Poon, M. Rockinger, and J. Tawn. New extreme-value dependence measures and finance applications. University of Strathclyde, 100 Cathedral Street, Glasgow G40LN, UK, February 2001.

[31] R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel, second edition, 2001.

[32] S. Resnick. Tail equivalence and its applications. *J. Applied Probability*, 8:136–156, 1971.

[33] S. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. Web available at http://www.orie.cornell.edu/~sid, 2001.

[34] S. Resnick. Modeling data networks. In *SemStat: Seminaire Europeen de Statistique, Exteme Value Theory and Applications*. Chapman-Hall, London, 2001. 75 pages; to appear.

[35] S. Resnick. On the foundation of multivariate heavy tail analysis. Submitted for C.C. Heyde festschrift; Web available at http://www.orie.cornell.edu/~sid, 2002.

[36] S. Resnick and P. Greenwood. A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis*, 9:206–221, 1979.

[37] S. Resnick, G. Samorodnitsky, and F. Xue. How misleading can sample acf's of stable ma's be? (Very!). *Ann.Applied Probability*, 9(3):797–817, 1999.

[38] S.I. Resnick. Point processes, regular variation and weak convergence. *Adv. Applied Probability*, 18:66–138, 1986.

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

[40] J. Rosiński. On the structure of stationary stable processes. *Ann. Probab.*, 23:1163–1187, 1995.

[41] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling. New York, NY: Chapman & Hall. xviii, 632, 1994.

[42] Ashoke Kumar Sinha. *Estimating Failure Probability when Failure is Rare: Multidimensional Case*. PhD thesis, Tinbergen Institute, Erasmus University Rotterdam, 1997. PhD Thesis # 165, Promotor: Prof. Dr. L.F.M. de Haan, published by Thela Thesis, Amsterdam.

[43] C. Stărică. Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance*, 6:515–553, 1999.

[44] C. Stărică. Multivariate extremes for models with constant conditional correlations. In P. Embrechts, editor, *Extremes and Integrated Risk Management*. Risk Books, London, 2000.

[45] J. Tawn. Bivariate extreme value theory: Models and estimation. *Biometrica*, 75:397–415, 1988.

[46] Yin Zhang, Lee Breslau, Vern Paxson, and Scott Shenker. On the characteristics and origins of internet flow rates. ACM Sigcom 2002 Conference, Pittsburgh, Pa; August 19-23, 2002.

Sidney Resnick, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853

www.orie.cornell.edu/~sid

E-mail address: sid@orie.cornell.edu