Asymptotic independence and support detection techniques from heavy-tailed multivariate data

Lehtomaa, Jaakko Resnick, Sidney^{*}

Abstract

One of the central objectives of modern risk management is to find a set of risks where the probability of multiple simultaneous catastrophic events is negligible. That is, risks are taken only when their joint behavior seems sufficiently independent. This paper aims to help to identify asymptotically independent risks by providing additional tools for describing dependence structures of multiple risks when the individual risks can obtain very large values.

The study is performed in the setting of multivariate regular variation. We show how asymptotic independence is connected to properties of the support of the angular measure. Secondly, we present an asymptotically consistent estimator of the support. The estimator generalizes to any dimension $N \ge 2$ and requires no prior knowledge of the support. The validity of the support estimate can be rigorously tested under mild assumptions by an asymptotically normal test statistic.

2010 MSC: Primary 62E20 60G70, Secondary 62G05 60G57

1 Introduction

This paper aims to contribute in answering to two questions: How can one decide if random variables are asymptotically independent and how can their dependence structure be analyzed in practice? In our approach the data is first thresholded based on the magnitudes of sample vectors and then divided into two parts. The first part is used to establish a grid based approximation of the asymptotic support. The remaining data is used to test the validity of the support estimate using an asymptotically normal test statistic.

^{*}Sidney Resnick was partially supported by US ARO MURI grant W911NF-12-1-0385.

In our view, the applied fields of finance and insurance, in particular, could benefit from having more robust tools for understanding extremal dependence. In finance, one of the central tasks of risk management is to identity which assets do not depend on others. The joint behavior of individual assets affects portfolio allocation strategies and ultimately determines which equities are chosen to a portfolio. One approach for reducing total risk is to select equities to be as independent as possible. In this way it is unlikely to experience many large shocks at once.

In insurance, large claims are a source of major concern because they can cause insolvency. The concern may become even more serious if there is a reasonable possibility of multiple lines of business accumulating large claims at the same time. Furthermore, successful pricing of insurance contracts as well as negotiations with reinsurance providers depend on having solid understanding of the worst case risks. For this reason, it is necessary to have an accurate observation based dependence model especially for the large claims.

We assume that all of the risks are *heavy-tailed*. A real valued random variable X is (right) heavy-tailed if it has no finite exponential moments, that is $\mathbb{E}(e^{sX}) = \infty$ for all s > 0. More precisely, we focus on a subclass of heavy-tailed distributions where the marginal distributions are *regularly varying*. We work under the assumptions that the sample is coming from a *multivariate regularly varying*, in short *MRV*, source. Properties of regularly varying and MRV distributions are comprehensively discussed in [22].

The problem of modeling extremal dependence structures has been studied in the past using various techniques. In the bivariate case it is common to use copulas and the index of tail dependence to quantify asymptotic dependence [9, 19]. The study of asymptotic dependence structures using MRVs is emphasised in particular in [2, 5, 6, 14, 17, 24]. Some approaches concentrate on modeling even finer properties such as the second order behavior of the limiting measure [3, 4, 16, 20]. Beyond the MRV setting, similar topics have been discussed from more extreme value theoretical viewpoint, see e.g. [8, Section 6] and its ample references. Finally, there exists an increasing number of publications which emphasise using the knowledge of dependence structures for the purposes of dimension reduction techniques [7, 10, 12]. One of the key differences between our method and most of the previous research is that we concentrate on approximating the asymptotic support, not the limiting measure. This goal seems to be more easily achievable with real data sets.

The core message of this paper is that there are applications in which some of the conventional risk measures are not adequate. Typically, the dependence structure of large observations determines the risk in the worst case scenario. In contrast, small observations might not have a significant impact even if they were highly dependent. Consider the following example to see why the most widely used risk statistic, correlation, might lead to counter-intuitive results.

Example 1.1. Let $\alpha > 2$ and l > 1. Suppose X, Z and B are independent random variables such that $\mathbb{P}(X > x) = \mathbb{P}(Z > x) = x^{-\alpha}$ for $x \ge 1$ and 1 otherwise. Let $\mathbb{P}(B = 0) = 1 - \mathbb{P}(B = 1) = 1/2$. Set

$$Y_1 := X\mathbf{1}(X \le l) + lZB\mathbf{1}(X > l)$$

and

$$Y_2 := Z \mathbf{1} (X \le l, Z \le l) + X \mathbf{1} (X > l).$$

Suppose the pairs (X, Z) and (X, Y_i) , i = 1, 2, denote risks to a company where the components of vectors correspond to different lines of businesses. The aim of the company is to avoid insolvency. From this viewpoint the pair (X, Y_1) should not be considered more risky than (X, Z) because the probability of two simultaneous catastrophic realizations is lower. On the other hand, the pair (X, Y_2) is much more risky than (X, Y_1) or (X, Z), because $Y_2 = X$ when X > l, resulting in two catastrophic realizations for sure when one line of business suffers a large loss. However, one would reach contradicting conclusions if correlation was used to quantify riskiness.

More precicely, $\operatorname{Corr}(X, Z) = 0$ due to independence of X and Z. However, $\operatorname{Corr}(X, Y_1) \to 1$, as $l \to \infty$. In addition, the pair (X, Y_2) is asymptotically fully dependent for all l > 1 using the terminology of [4]. Yet, $\operatorname{Corr}(X, Y_2) \to 0$, as $l \to \infty$. So, using correlation as a measure of risk in this example leads to overestimation of insignificant risk as well as underestimation of potentially catastrophic risk.

The failure of risk quantification in Example 1.1 is due to the fact that correlation, among other popular risk metrics, has only a limited capability of describing dependence of rare events. Similar phenomena as in the artificial Example 1.1 have been observed in nature. In [23], the authors study meteorological data in order to model extreme ground level ozone events. The study depicts cases where the etremal observations have significantly different dependence structure than small observations, see e.g. Figure 1 of [23].

In conclusion, modeling dependence structures with emphasis on accuracy of tail behavior requires different tools than modeling systems as a

whole. The methods presented in Sections 2 and 3 cover some of the shortcomings of previous approaches. In particular, they allow practitioners to verify which risks fulfill the goals of risk management in finance and insurance when the MRV framework is applicable.

1.1 Structure of the paper

The rest of Section 1 is used to define concepts and definitions. In Section 2, the grid based asymptotic support estimator for multivariate heavy-tailed data is presented. Consistency and related properties are proved in Section 2.2. The definition of asymptotic independence as well as the new connections with limiting behavior of MRVs are discussed in Section 3. Specifically, the test for asymptotic normality is introduced in Section 3.3. In Section 4, the techniques developed in Sections 2-3 are illustrated by means of simulated and real examples.

1.2 Basic definitions

Suppose $(\Omega, \mathcal{B}, \mathbb{P})$ is a probability space where all the subsequent random variables are defined. Throughout the paper random variables take values in a metric space $(\mathbb{R}^N, \mathcal{T}, d)$. Here $N \geq 2$ is the dimension of the space, \mathcal{T} is the Euclidean topology and $d = d_{\mathbb{R}^N}$ is the L_2 or Euclidean distance. That is, for all elements $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, where $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(N)})$ and $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(N)})$,

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{N} (x^{(i)} - y^{(i)})^2}.$$

Euclidean distances are used in mappings that project sets into lower dimensional spaces in a way that does not distort the image. However, unless otherwise stated, all norms denoted by $|| \cdot ||$ are L_1 -norms, where

$$||\mathbf{x}|| = \sum_{i=1}^{N} |x^{(i)}|.$$

The choice of L_1 norm instead of some other L_p norm is natural because in applications the total risk is typically the sum of marginal risks. So, any condition on the size of the L_1 norm can be directly viewed as a condition on the total risk. Upper indices are used to identify components of vectors. Lower indices are reserved for order statistics. For $1 \le i \le N$, the *i*th largest component of **x** is $x_{(i)}$. All inequalities and operations involving vectors are understood componentwise way as in Section 1.2 of [20].

The collections of open, closed and compact sets are denoted by \mathcal{G},\mathcal{F} and \mathcal{K} , respectively. For a set $A \subset \mathbb{R}^N$ the whole space can be partitioned as $\mathbb{R}^N = \operatorname{int}(A) \cup \operatorname{ext}(A) \cup \partial A$ to topological interior, exterior and boundary of the set A. The ball with center $\mathbf{x} \in \mathbb{R}^N$ and radius $\delta > 0$ is $B(\mathbf{x}, \delta)$. In addition, the diameter of A is denoted by diam(A). The notation := is used when the left hand side is defined by the right hand side of the equation.

1.3 Multivariate regular variation

We follow the standard definition of multivariate regular variation as defined in [22, Theorem 6.1]. In our case, however, the definition is slightly modified to take into account possible negative values of components. Note that normalizing all components using the same function b implies that the components must be tail equivalent, see [22, Remark 6.1.].

Definition 1.1. Suppose $\mathbf{Z} = (Z^{(1)}, Z^{(2)}, \dots, Z^{(N)})$ is an random vector in \mathbb{R}^N . Set $\mathbb{E} := [-\infty, \infty]^N \setminus \{\mathbf{0}\}$. We say that \mathbf{Z} is standard multivariate regularly varying with limit measure ν if there exist a function $b(t) \uparrow \infty$, as $t \to \infty$, such that

(1.1)
$$t\mathbb{P}\left(\frac{\mathbf{Z}}{b(t)}\in\cdot\right)\left(\stackrel{v}{\leftrightarrow}\nu\right)$$

in $M_+(\mathbb{E})$. Notation \xrightarrow{v} stands for vague convergence of measures.

Multivariate regular variation has an equivalent definition via measure S, called the *angular measure*, *spectral measure* or *limiting measure* in different sources, defined on

(1.2)
$$C^N := \{ \mathbf{z} \in \mathbb{R}^N : ||\mathbf{z}|| = 1 \}.$$

In this formulation, **Z** is said to be standard multivariate regularly varying if there exist a function $b(t) \uparrow \infty$, as $t \to \infty$, such that for

$$(R,\Theta) := \left(||\mathbf{Z}||, \frac{\mathbf{Z}}{||\mathbf{Z}||} \right) \left($$

we have

(1.3)
$$t\mathbb{P}\left(\left(\frac{R}{\mathbf{b}(t)},\Theta\right)\in\cdot\right)\overset{v}{\longleftrightarrow}c\nu_{\alpha}\times\mathbb{S}$$

in $M_+((0,\infty] \times C^N)$, as $t \to \infty$, where c > 0, \mathbb{S} is a probability measure on C^N and $\nu_{\alpha}((x,\infty]) = x^{-\alpha}$. The number $\alpha > 0$ is called the *tail index* of the multivariate regularly varying distribution.

1.4 *N*-simplex and simplex mappings

In Section 2, the aim is to identify the support of the limiting measure S based on data. To this end, we present a support estimation method in which the support of S on L_1 sphere is approximated by a set consisting of equally sized rectangles. The locations of rectangles are determined based on concentrations of probability mass of the empirical version of the limiting measure S. This is in contrast to [4], where the range of thresholded data itself is used to indicate the location of the support. However, to our experience in working with data sets, finding the sets of highest concentration provide a way to eliminate noise arising from unlikely observations that lie outside of the asymptotic support.

Suppose $m \ge 2$ is an integer and $N \ge 2$ is the dimension of the data. Now, m determines the *resolution* of the asymptotic support estimate. The idea is to map the N-dimensional simplex into $[0, 1]^{N-1}$ one face at a time. The image on is partitioned into m^{N-1} smaller sets. The partition is called a grid and the sub squares are called *cells*. Some of the grid cells are accepted as part of the support while the rest are rejected based on a rule described in Section 2.

In addition to the N-simplex in L_1 , we set

$$C_{+}^{N} := \{ \mathbf{z} \in \mathbb{R}_{+}^{N} : z^{(1)} + z^{(2)} + \ldots + z^{(N)} = 1 \}$$

to denote the part of simplex C^N where all coordinates are non-negative. If $A \subset \{1, 2, ..., N\}$ is a set of indices we define the *faces* by formula

$$C^{N}(A) := \{ \mathbf{z} \in C^{N} : z^{(i)} = 0, \text{ when } i \notin A \}.$$

Definition 1.2 (Support of measure in \mathbb{R}^N). Let $E = (\mathbb{R}^N, \mathcal{T}, d)$ be a topological space where \mathcal{T} is the smallest sigma algebra containing all open balls in metric d which is the Euclidean metric. Suppose μ is a measure on $(\mathbb{R}^N, \mathcal{T})$.

Then the support $\operatorname{supp}(\mu)$ of measure μ is the set defined as

(1.4)
$$\operatorname{supp}(\mu) := \left\{ x \in \mathbb{R}^N : \mu(B(x,\delta)) > 0 \text{ for all } \delta > 0 \right\}$$

Equivalently, it is the complement of the union of open balls with measure zero. In particular, $supp(\mu)$ is the smallest closed set in the sense that

$$\operatorname{supp}(\mu) = \bigcap_{\substack{A \in \mathcal{F} \\ \mu(A^c) = 0}} A.$$

The part of the support of S on simplex C^N_+ is denoted by \mathbf{supp}^+ .

Definition 1.3. Let $N \geq 2$. Suppose T is a bijective mapping $T: C^N_+ \to [0,1]^{N-1}$ with property

$$d_{\mathbb{R}^N}(\mathbf{x}, \mathbf{y}) = a d_{\mathbb{R}^{N-1}}(T(\mathbf{x}), T(\mathbf{y}))$$

for all $\mathbf{x}, \mathbf{y} \in C^N_+$ and some constant a > 0. Such a mapping T is called *simplex mapping* associated with C^N_+ .

Mapping T can be chosen in a number of ways. So, the grid positioning can be adjusted with respect to observed data if necessary. By shifting the grid one can avoid concentration of points directly onto a grid boundaries. The positive simplex is mapped into a lower dimensional space for clarity. Specifically, when N = 3, all analysis can be performed in the two dimensional plane.

Example 1.2. a) If N = 2, one can set

$$T\left(\begin{bmatrix} z_1\\ z_2\end{bmatrix}\right) \rightleftharpoons z_1.$$

b) If N = 3, setting

$$T\left(\begin{bmatrix}z_1\\z_2\\z_3\\z_3\end{bmatrix}\right) = \begin{bmatrix}\frac{1}{2}(z_2-z_1+1)\\\frac{\sqrt{3}}{2}z_3\end{bmatrix}$$

gives T that maps C^3_+ into $[0,1]^2$. The image $T(C^3_+)$ is a region in $[0,1]^2$ inside an equilateral triangle with edges on (0,0), (1,0) and $(1/2,\sqrt{3}/2)$. Mapping T has an inverse $T^{-1}: T(C^3_+) \to C^3_+$ given by

$$T^{-1}\left(\begin{bmatrix} \widetilde{z}_1\\ \widetilde{z}_2\end{bmatrix}\right) \left(= \begin{bmatrix} 1 \left(-z_1 - \frac{z_2}{\sqrt{3}}\right)\\ z_1 - \frac{z_2}{\sqrt{3}}\\ \frac{2z_2}{\sqrt{3}}\end{bmatrix}\right)$$

Mappings of Example 1.2 can be used in computer programs that require explicit formulas for projections.

2 Support estimation

In this section the grid based estimator is defined and its asymptotic consistency is shown under general assumptions.

2.1 Support estimator and related quantities

Suppose **Z** is a multivariate regularly varying vector in \mathbb{R}^N . Let $\mathbf{s} \in \{-1, 1\}^N$ be a vector. We define

(2.1)
$$\mathbf{U} := T\left(\frac{\mathbf{Z}}{||\mathbf{Z}||}\right) \left($$

and more generally,

$$\mathbf{U}(\mathbf{s}) := T \quad \frac{(s^{(i)}Z^{(i)})_{i=1}^{N}}{||\mathbf{Z}||} \Bigg) \Bigg($$

to denote the transformed angular component of facet \mathbf{s} in $T(C_+^N) \subset [0, 1]^{N-1}$. Most of the proofs can be formulated in terms of \mathbf{U} , which corresponds to the positive facet. For example if N = 3, then $\mathbf{U} = \mathbf{U}((1, 1, 1))$.

The space $[0,1]^{N-1}$ can be covered by separate cells. Given a vector $\mathbf{x} \in [0,1]^{N-1}$ and a number $m \geq 2$, we define a cell $M(\mathbf{x},m) \subset \mathbb{R}^{N-1}$ by formula

(2.2)
$$M(\mathbf{x},m) := \mathbf{x} + [0,1/m)^{N-1}$$

Cell $M(\mathbf{x}, m)$ can be viewed as a shift by vector \mathbf{x} . Explicitly, $M(\mathbf{x}, m)$ is the set

$$\left[x^{(1)}, x^{(1)} + \frac{1}{m}\right) \times \left[x^{(2)}, x^{(2)} + \frac{1}{m}\right) \times \ldots \times \left[x^{(N-1)}, x^{(N-1)} + \frac{1}{m}\right).$$

The aim is to rasterize projected observations in order to produce an estimate for the asymptotic support. This is done by partitioning the set $T(C_+^N)$ by grid cells.

Definition 2.1 (Support estimator). Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$ be i.i.d. multivariate regularly varying vectors. Suppose k and m are natural numbers such that $k \geq 1$ and $m \geq 2$. Define G_m to be the set of corner points of cells at resolution m by

$$G_m := \left\{ \mathbf{x} \in \mathbb{R}^{N-1} : x^{(i)} \in \left\{ \emptyset, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}, 1 \right\}, i = 1, 2, \dots, N-1 \right\} \left(\left(\sum_{i=1}^{n} \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}, 1 \right) \right\}$$

For $q \in [0, 1]$ and facet $\mathbf{s} \in \{-1, 1\}^N$, support estimator $A(\mathbf{s}) = A_{k,m,q}(\mathbf{s})$ is a set defined as

$$A(\mathbf{s}) := \bigcup_{\mathbf{x} \in G_m} \left\{ \mathbf{M}(\mathbf{x}, m) : \frac{\sum_{i=1}^m \mathbf{1}(\mathbf{U}_i(\mathbf{s}) \in \mathbf{M}(\mathbf{x}, m), ||\mathbf{Z}_i|| \ge ||\mathbf{Z}_{(k)}||)}{k} > q \right\}.$$

The estimator corresponding to the positive facet $s = \mathbf{1} = (1, 1, ..., 1)$ is denoted by $A_{k,m,q} := A_{k,m,q}(\mathbf{1})$.

Support estimator $A_{k,m,q}$ is a random set formed based on a random sample $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$. It has three parameters: k, m and q. Parameter k = k(n) is the number of order statistics used from the sample. For the asymptotic analysis we assume that $n/k(n) \to \infty$, as $n \to \infty$. Parameter mdenotes the resolution at which the estimate is formed. In asymptotic results resolution grows so that cell size decreases. One can think of parameter q as a rejection threshold. It determines how many observations are needed in a single grid cell for the cell to be accepted as part of the support estimate. In practice it helps to reject unlikely observations and noise. If p observations are required in a given sample of n one can set q = p/k(n).

Support estimators in Definition 2.1 are decreasing in q. For fixed k and m the inclusion $A_{k,m,q_2} \subset A_{k,m,q_1}$ holds for $0 < q_1 < q_2 < 1$. Furthermore, limiting behavior as $m \to \infty$ can be studied in a sequence of nested grids by considering dyadic resolutions $m = 2^s$, where $s = 1, 2, \ldots$ However, in applications the geometric convergence of cell diameters may turn out to be too fast which is why m is allowed to be any positive integer, not only a power of, say, number 2.

Definition 2.2. (Rasterized support) The smallest grid set with resolution m that contains $supp^+$ is called the *rasterised support* and defined by

$$\operatorname{supp}_{m}^{+} := \bigcup \left\{ \operatorname{M}(\mathbf{x}, m) : \mathbf{x} \in G_{m}, \, \mathbb{S}(\operatorname{M}(\mathbf{x}, m)) > 0 \right\}.$$

2.2 Consistency of the grid based support estimator

Lemma 2.1. Suppose $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \ldots$ are *i.i.d* random vectors with a common multivariate regularly varying distribution. Assume further that $n \to \infty$ and $n/k \to \infty$. Recall from ... that G_m is the grid corresponding to resolution m.

If
$$\mathbb{S}_T(G_m) = 0$$
, then

(2.3)
$$\frac{\#\left\{i \le n : U_i \in \mathbf{m}(m_1, m_2), ||\mathbf{Z}_i|| \ge ||\mathbf{Z}_{(k)}||\right\}}{k} \left(\xrightarrow{P} \mathbb{S}_T(\mathbf{m}(m_1, m_2)), \right)$$

as $n \to \infty$ for all sets $\mathbf{m}(m_1, m_2)$, where $1 \le m_1, m_2 \le m$.

Proof. (Sketch of Proof) Suppose $1 \leq m_1, m_2 \leq m$ and denote $A := T^{-1}(\mathbf{m}(m_1, m_2))$. The set A is a continuity set of measure \mathbb{S} by assumption $\mathbb{S}_T(G_m) = 0$, that is, $\mathbb{S}(\partial A) = 0$. Note that the left hand side of (2.3) can be written as

$$\frac{1}{k}\sum_{i=1}^{n} \mathbf{1}(||\mathbf{Z}_{i}|| > ||\mathbf{Z}_{(k+1)}||, \mathbf{Z}_{i} > ||\mathbf{Z}_{(k+1)}|| \in A).$$

Proposition 6.2 of [2, p. 158] states that in this case

(2.4)
$$\hat{S}_{n}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}(||\mathbf{Z}_{i}|| > ||\mathbf{Z}_{(k+1)}||) \epsilon_{\mathbf{Z}_{i}/||\mathbf{Z}_{i}||}(\cdot) \Rightarrow \mathbb{S}$$

in $\mathbb{P}(C_2)$, the space of probability measures on C_2 . This implies

$$\hat{S}_n(A) \Rightarrow \mathbb{S}(A),$$

as $n \to \infty$. The distributional limit is a constant, which is why convergence takes place also in probability and (2.3) holds.

Proposition 2.1 (Consistency of the grid estimator). Suppose assumptions of Lemma 2.1 hold. Let k = k(n) be such that $k(n) \to \infty$ and $n/k(n) \to \infty$, as $n \to \infty$. Assume further that $m \ge 2$ and

(2.5)
$$q \in (0,1) \setminus \{ \mathbb{S}_T(\mathbf{m}(m_1, m_2)) : 1 \le m_1 \le m, 1 \le m_2 \le m \}.$$

Let $A_{k,m,q}$ be defined as in Definition 2.1.

Then, it holds for a fixed pair (m_1, m_2) that

(2.6)
$$\mathbb{P}\left(A_{k,m,q}(m_1,m_2) = \operatorname{supp}_{m,q}^+(m_1,m_2)\right) \longleftrightarrow 1,$$

as $n \to \infty$.

Proof. Suppose first that $\mathbb{S}_T(\mathbf{m}(m_1, m_2)) > q$ so that $\mathbf{supp}_{m,q}^+(m_1, m_2) = 1$. Then the probability in Equation (2.6) can be written as

$$\mathbb{P}\left(\frac{\#\left\{i\left(:U_i\in\mathbf{m}(m_1,m_2),||\mathbf{Z}_i||\geq||\mathbf{Z}_{(k)}||\right\}}{k}>q\right)\left(\frac{1}{k}\right)$$

By Lemma 2.1,

$$\frac{\#\left\{i: U_i \in \mathbf{m}(m_1, m_2), ||\mathbf{Z}_i|| \ge ||\mathbf{Z}_{(k)}||\right\}}{k} \left(\xrightarrow{P} \mathbb{S}_T(\mathbf{m}(m_1, m_2)) \right)$$

and $\mathbb{S}_T(\mathbf{m}(m_1, m_2)) > q$ by assumption. This shows (2.6).

If $\mathbb{S}_T(\mathbf{m}(m_1, m_2)) < q$ so that $\mathbf{supp}_{m,q}^+(m_1, m_2) = 0$, the proof is similar as in the first case, but the studied probability is

$$\mathbb{P}\left(\frac{\#\left\{i\left(:U_i\in\mathbf{m}(m_1,m_2),||\mathbf{Z}_i||\geq||\mathbf{Z}_{(k)}||\right\}}{k}\leq q\right)\right)\left(\frac{1}{k}\right)$$

Example 2.1. In Equation (2.5) of Theorem 2.1 finitely many values for q are excluded. This is necessary, because... two points, oscillation of probability mass around q in one of them.

Corollary 2.1. Suppose assumptions of Theorem 2.1 hold. Then

1. For fixed $m \ge 2$, support supp^+ is an eventual subset of the estimating grid in the sense that

(2.7)
$$\mathbb{P}\left(\operatorname{supp}^{+}\subset\bigcup_{q>0}(A_{k,m,q})\to 1,\right.$$

as $n \to \infty$.

2. For fixed m and q,

(2.8)
$$\mathbb{P}\left(\bigwedge_{k,m,q} = \mathbf{supp}_{m,q}^{+}\right) \leftrightarrow 1,$$

as $n \to \infty.$

Proof. Proof of Part 1: Denote

$$\hat{q} := \min\{\mathbb{S}_T(\mathbf{m}(m_1, m_2)) > 0 : (m_1, m_2) \in B_{\sup \mathbf{p}_m^+}\}$$

Since $\operatorname{supp}^+ \subset \operatorname{supp}_m^+$, it holds that

$$\begin{aligned}
\mathbb{P}(\mathbf{supp}^+ \subset \cup_{q>0} A_{k,m,q}) \\
\geq \mathbb{P}(\mathbf{supp}_m^+ \subset \cup_{q>0} A_{k,m,q}) \\
\geq \mathbb{P}(\mathbf{supp}_m^+ \subset A_{k,m,\hat{q}/2}) \\
(2.9) &= \mathbb{P}(A_{k,m,\hat{q}/2}(m_1,m_2) = 1 \text{ for all } (m_1,m_2) \in B_{\mathbf{supp}_m^+}).
\end{aligned}$$

Events in (2.9) are not independent, but each of them will have probability 1 in the limit $n \to \infty$ by Theorem 2.1. Since the number of events is at most m^2 and thus finite, the claim holds.

Proof of Part 2: Probability in (2.8) is

$$\mathbb{P}(A_{k,m,q}(m_1, m_2) = \mathbf{supp}_{m,q}^+ \text{ for all } 1 \le m_1 \le m, 1 \le m_2 \le m)$$

and consists of events that have probability 1 in the limit $n \to \infty$ similarly as in the proof of Part 1.

Proposition 2.2. Suppose assumptions of Theorem 2.1 hold. Then, for fixed m,

(2.10)
$$\mathbb{P}(A_{k(n),m} = \mathbf{supp}_m^+) \to 1,$$

as $n \to \infty$.

Sketch of proof. If $\operatorname{supp}^+(m_1, m_2) = 1$ the proof is similar to the first part of Theorem 2.1. So, in order to prove (2.10) it suffices to concentrate on the case where $\operatorname{supp}^+(m_1, m_2) = 0$ and show that

$$\mathbb{P}(A_{k,m}(m_1, m_2) = 0) \to 1,$$

as $n \to \infty$. To do this, one can estimate probability $\mathbb{P}(A_{k,m}(m_1, m_2) = 1)$ by an upper bound that converges to 0. Now, for a fixed n,

(2.11)
$$\mathbb{P}(A_{k,m}(m_1, m_2) = 1) = \mathbb{P}(\hat{S}_n(\mathbf{m}(m_1, m_2)) > 0),$$

where $\hat{S}(\cdot)$ is the empirical measure defined in (2.4). Since $\mathbb{S}_T(\mathbf{m}(m_1, m_2)) = 0$ and $\mathbf{m}(m_1, m_2)$ is a continuity set of \mathbb{S}_T by assumption, the quantity in (2.11) converges to 0, as $n \to \infty$.

Theorem 2.1 (Consistency). Suppose assumptions of Proposition 2.2 hold. In addition, assume that $m: \mathbb{N} \to \mathbb{N}$ is a function such that $m(n) \uparrow \infty$, as $n \to \infty$.

Then $A_{k(n),m(n)} \to \operatorname{supp}^+$ in the sense that

(2.12)
$$D(A_{k(n),m(n)}, \operatorname{supp}^+) \xrightarrow{P} 0,$$

as $n \to \infty$ where D is the Hausdorff distance. Equivalently, for any $\delta > 0$,

(2.13)
$$\mathbb{P}(A_{k(n),m(n)} \subset (\mathbf{supp}^+)^{\delta}) \to 1, \quad n \to \infty$$

and

(2.14)
$$\mathbb{P}(\mathbf{supp}^+ \subset A^{\delta}_{k(n),m(n)}) \to 1, \quad n \to \infty.$$

Proof. (Sketch of proof) Proof of Equation (2.13): Suppose $\delta > 0$ is fixed. Let m_{δ} be so large that diam $(\mathbf{m}_{\delta}(1,1)) < \delta/4$. Set

$$C_1 := \bigcup_{\substack{m_1, m_2:\\ \mathbf{m}_{\delta}(m_1, m_2) \setminus (\mathbf{supp}^+)^{\delta/2} \neq \emptyset}} \mathbf{m}_{\delta}(m_1, m_2).$$

Now C_1 is a collection of cells at resolution m_{δ} that do not touch $(\mathbf{supp}^+)^{\delta/2}$, but yet $((\mathbf{supp}^+)^{\delta})^c \subset C_1$.

Since diam $(\mathbf{m}_{\delta}(1,1)) < \delta/4$, it must hold that $d(C_1, \mathbf{supp}^+) > 0$. In particular, $C_1 \cap \mathbf{supp}^+ = \emptyset$, which implies $\mathbb{S}_T(C_1) = 0$ and $\mathbb{S}_T(\partial C_1) = 0$. The latter equality follows from the fact that C_1 is a finite union of cells that are assumed to be continuity sets of \mathbb{S}_T .

So, if n is so large that $\operatorname{diam}(\mathbf{m}(n)(1,1)) < \delta/4$, then

$$\{ \#\{i \le n : U_i \in C_1, ||\mathbf{Z}_i|| \ge ||\mathbf{Z}_{(k)}|| \} = 0 \quad \subset \{A_{k(n),m(n)} \subset (\mathbf{supp}^+)^{\delta} \}.$$

This implies (2.13) because $\mathbb{P}(A_{k(n),m(n)} \subset (\mathbf{supp}^+)^{\delta})$ is now bounded from below by a probability that converges to 1 by Proposition 2.2, as $n \to \infty$.

Proof of Equation (2.14): Suppose $\delta > 0$ is fixed. Let m_{δ} be so large that diam $(\mathbf{m}_{\delta}(1,1)) < \delta/2$. Define

$$C_2 := \bigcup_{\substack{m_1, m_1:\\ \inf(\mathbf{m}_{\delta}(m_1, m_2)) \\ \mathbf{Supp}^+ \neq \emptyset}} \mathbf{m}_{\delta}(m_1, m_2).$$

The set C_2 is the collection of cells at resolution m_{δ} that contains all probability mass of the limit measure \mathbb{S}_T . This follows from the fact that all cells are assumed to be continuity sets of \mathbb{S}_T .

Note first that the δ -swelling of even a single point in a cell of C_2 contains the cell itself due to assumption diam $(\mathbf{m}_{\delta}(1,1)) < \delta/2$. Suppose then that there is at least one observation in each cell of C_2 . In this case $\mathbf{supp}^+ \subset A_{k(n),m(n)}^{\delta}$. This is due to the fact that the possible boundary points in $\mathbf{supp}^+ \setminus C_2$ must be boundary points of some cell that does belong to C_2 and thus are included in the δ -swelling. In conclusion, one way in which the inclusion $\mathbf{supp}^+ \subset A_{k(n),m(n)}^{\delta}$ can hold is that there is at least one observation in each cell of C_2 . So,

$$\{A_{k(n),m_{\delta}}(m_1,m_2)=1 \text{ for all cells in } C_2 \subset \{\mathbf{supp}^+ \subset A_{k(n),m(n)}^{\delta}\}.$$

This implies (2.14) because $\mathbb{P}(\mathbf{supp}^+ \subset A_{k(n),m(n)}^{\delta})$ is now bounded from below by a probability that converges to 1 by Proposition 2.2, as $n \to \infty$. \Box

Remark 2.1. Convergence in the sense of Theorem 2.1 does not guarantee that the approximation covers the support. In fact, if m grows rapidly with respect to k, the approximation may have zero Lebesgue measure. One would need to set bounds for the growth of m in order to get such result.

3 Asymptotic independence

The concept of asymptotic independence is wider than independence. If a random vector has asymptotically independent components, one large component of the vector reveals no information from the other components. It admits all dependence structures as long as they cannot produce realizations of vectors where multiple components obtain large values at the same time. From practical viewpoint asymptotically independent components are as harmless as independent components. So, omitting asymptotically independent components from analysis is a way to reduce dimension of a studied system. Doing so increases the accuracy of estimates of asymptotic support when only a fixed amount of data is available. The topic of dimension reduction in models with extremal dependence is also discussed in [10, 24].

Our definition of asymptotic independence is compatible with existing litterature. In particular, it follows the definition given in [22, p 195]. The new property is that several groups of components can be handled at once. The closest results in this field are, to our knowledge, the method of extremograms discussed in e.g. [14,17] and the sparcity approach to dimension reduction presented in [10].

Definition 3.1 below would not be suitable if the marginals were not heavy-tailed. The behavior of vectors composed of sufficiently light-tailed i.i.d. components would be different. In fact, the conditional probability distribution $\mathbb{S}_{T,l}$ might concentrate on the centers of simplex faces, as $l \to \infty$. This is in contrast to the i.i.d. heavy-tailed setting where the concentration would take place at the extremal points of the simplex. See [15] for details in the two dimensional case.

3.1 Definition of asymptotic independence of MRV

Definition 3.1. [Asymptotic independence for MRV] Suppose $\mathbf{Z} \geq \mathbf{0}$ has regularly varying multivariate distribution with scaling function b. Let $A_1, A_2 \subset \{1, 2, \ldots, N\}$ and suppose $\#A_1 = N_1$ and $\#A_2 = N_2$. We say that component $\mathbf{Z}_{A_1} := (Z^{(i)})_{i \in A_1}$ is asymptotically independent of component $\mathbf{Z}_{A_2} := (Z^{(i)})_{i \in A_2}$ if

(3.1)
$$t\mathbb{P}\left(\frac{\mathbf{Z}_{A_1}}{b(t)} \in B_1, \frac{\mathbf{Z}_{A_2}}{b(t)} \in B_2\right) \longleftrightarrow 0$$

holds for all $B_1 \subset \mathbb{R}^{N_1}$ and $B_2 \subset \mathbb{R}^{N_2}$ such that $d(B_1, \mathbf{0}) > 0$ and $d(B_2, \mathbf{0}) > 0$, as $t \to \infty$.

Remark 3.1. It may be assumed without loss of generality that the sets B_1 and B_2 in (3.1) are N_1 and N_2 dimensional rectangles. The statement is made more precise in Part 1 of Theorem 3.1.

Next, we define projections and methods that can be used to combine multiple components of random vectors into a single group. It enables the study of two groups in a simple setting even though the original data set is high dimensional.

Definition 3.2. Let $A_1, A_2 \subset \{1, 2, ..., N\}$, $A_1 \cap A_2 = \emptyset$ and suppose $\#A_1 = N_1$ and $\#A_2 = N_2$, where $N_1, N_2 \ge 1$ and $N_1 + N_2 = N$. Define vectors $\mathbf{a}_1, \mathbf{a}_2 \in C_+^N$ by formulas

$$\mathbf{a}_1^{(i)} = \begin{cases} 1/N_1, & i \in A_1 \\ 0, & i \notin A_1 \end{cases}$$

and

$$\mathbf{a}_2^{(i)} = \begin{cases} 1/N_2, & i \in A_2\\ 0, & i \notin A_2. \end{cases}$$

Vectors \mathbf{a}_1 and \mathbf{a}_2 are called the *midpoints of faces* $C^N(A_1)$ and $C^N(A_2)$, respectively.

Midpoints \mathbf{a}_1 and \mathbf{a}_2 are linearly independent vectors in \mathbb{R}^N . For this reason the subspace $W_{\mathbf{a}_1,\mathbf{a}_2} := \operatorname{span}(\mathbf{a}_1,\mathbf{a}_2)$ spanned by the midpoints is a plane. We can thus define orthogonal projections onto the subspace $W_{\mathbf{a}_1,\mathbf{a}_2}$ via projection matrix $Q_{\mathbf{a}_1,\mathbf{a}_2} := \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$, where \mathbf{M} is the $N \times 2$ matrix $\mathbf{M} = [\mathbf{a}_1,\mathbf{a}_2]$.

In our case, where the subspace is spanned by midpoints, the projection matrix Q_{a_1,a_2} is of particularly simple form. It can be seen by a direct calculation that

$$(3.2) \qquad \qquad \boldsymbol{Q}_{\mathbf{a}_1,\mathbf{a}_2} = [\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_N],$$

where

$$\mathbf{c}_i = \begin{cases} \mathbf{a}_1', & \mathbf{a}_1^{(i)} \neq 0\\ \mathbf{a}_2', & \mathbf{a}_2^{(i)} \neq 0. \end{cases}$$

Example 3.1. Suppose N = 5, $A_1 = \{1, 2, 4\}$ and $A_2 = \{3, 5\}$. Now

 $\mathbf{a}_1 = [\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3}, 0]^T, \, \mathbf{a}_2 = [0, 0, \frac{1}{2}, 0, \frac{1}{2}]^T$ and

Orthogonally projected points are connected to linear combinations of midpoints \mathbf{a}_1 and \mathbf{a}_2 . An orthogonally projected point $\mathbf{x} \in \mathbb{R}_+$ has presentation

(3.3)
$$\boldsymbol{Q}_{\mathbf{a}_1,\mathbf{a}_2}\mathbf{x} = \left(\sum_{i\in A_1} x^{(i)}\right) \mathbf{a}_1 + \left(\sum_{i\in A_2} x^{(i)}\right) \mathbf{a}_2.$$

Next, we will define projections that allow projection of multidimensional data onto a line. The projected points can be used to inspect validity of asymptotic independence.

Definition 3.3. Let A_1 and A_2 be as in Definition 3.2 and $Q_{\mathbf{a}_1,\mathbf{a}_2}$ as in (3.2).

Mappings $h_1: \mathbb{R}^N_+ \setminus \{\mathbf{0}\} \to C^N_+$, $h_2: \mathbb{R}^N_+ \to \mathbb{R}^N_+$ and $h_3: \{(1-t)\mathbf{a}_1 + t\mathbf{a}_2 : t \in [0,1]\} \to [0,1]$ are defined as

$$h_1(\mathbf{x}) := \frac{\mathbf{x}}{||\mathbf{x}||},$$
$$h_2(\mathbf{x}) := \boldsymbol{Q}_{\mathbf{a}_1, \mathbf{a}_2} \mathbf{x}$$

and

$$h_3(\mathbf{x}) := h_4^{-1}(\mathbf{x})$$

where h_4 is the linear interpolation $h_4(t) = (1-t)\mathbf{a}_1 + t\mathbf{a}_2, t \in [0,1]$. We define projection $\operatorname{proj}_{\mathbf{a}_1,\mathbf{a}_2} \colon \mathbb{R}^N_+ \setminus \{\mathbf{0}\} \to [0,1]$ by

(3.4)
$$\operatorname{proj}_{\mathbf{a}_1,\mathbf{a}_2}(\mathbf{x}) := h_3(h_2(h_1(\mathbf{x}))).$$

Function $\operatorname{proj}_{\mathbf{a}_1,\mathbf{a}_2}(\mathbf{x})$ projects points of $\mathbb{R}_+ \setminus \{\mathbf{0}\}$ first onto L_1 simplex and then orthogonally onto the line connecting midpoints \mathbf{a}_1 and \mathbf{a}_2 . The order of projections h_1 and h_2 can be switched.

Lemma 3.1. Suppose $\mathbf{x} \in \mathbb{R}_+ \setminus \{\mathbf{0}\}$. Let $A_1 A_2$, h_1 and h_2 be as in Definition 3.3.

Then

(3.5)
$$h_2(h_1(\mathbf{x})) = h_1(h_2(\mathbf{x})).$$

Proof. We note first that $Q_{\mathbf{a}_1,\mathbf{a}_2}\mathbf{x} \in \mathbb{R}_+ \setminus \{\mathbf{0}\}$ so that the function $h_1(h_2(\mathbf{x}))$ is well defined. Furthermore, because $Q_{\mathbf{a}_1,\mathbf{a}_2} = Q_{\mathbf{a}_1,\mathbf{a}_2}^T$, it holds that

(3.6)
$$||\mathbf{Q}_{\mathbf{a}_1,\mathbf{a}_2}\mathbf{x}|| = \sum_{i=1}^{N_1} \mathbf{a}_1 \cdot \mathbf{x} + \sum_{i=1}^{N_2} \mathbf{a}_2 \cdot \mathbf{x} = \sum_{i=1}^{N} \mathbf{x}^{(i)} = ||\mathbf{x}||.$$

Now, using linearity of h_2 and Equation (3.6) we get

$$h_2(h_1(\mathbf{x})) = \mathbf{Q}_{\mathbf{a}_1, \mathbf{a}_2} \frac{\mathbf{x}}{||\mathbf{x}||} = \frac{\mathbf{Q}_{\mathbf{a}_1, \mathbf{a}_2} \mathbf{x}}{||\mathbf{x}||} = \frac{\mathbf{Q}_{\mathbf{a}_1, \mathbf{a}_2} \mathbf{x}}{||\mathbf{Q}_{\mathbf{a}_1, \mathbf{a}_2} \mathbf{x}||} = h_1(h_2(\mathbf{x})).$$

Lemma 3.1 states that the mapping $\operatorname{proj}_{\mathbf{a}_1,\mathbf{a}_2}$ of Definition 3.3 can be viewed in two different ways. This is relevant in the proof of the following result.

3.2 Connection between asymptotic independence and the limit measure

Theorem 3.1. Suppose $\mathbf{Z} \geq \mathbf{0}$ is a multivariate regularly varying random vector. Let \mathbf{Z} , \mathbf{Z}_{A_1} and \mathbf{Z}_{A_2} be as in Definition 3.1 and $A_1 \cap A_2 = \emptyset$. Then the following are equivalent with (3.1):

1) Suppose $B_1 \subset \mathbb{R}^N$ and $B_2 \subset \mathbb{R}^N$ are Borel sets bounded away from **0**. Assume further that the sets B_1 and B_2 can be presented as

$$B_1 = B_1^{(1)} \times B_1^{(2)} \times \ldots \times B_1^{(N)}$$
, where $B_1^{(i)} = \mathbb{R}$ for all $i \in A_2$

and

$$B_2 = B_2^{(1)} \times B_2^{(2)} \times \ldots \times B_2^{(N)}$$
, where $B_2^{(i)} = \mathbb{R}$ for all $i \in A_1$.

Then

$$t\mathbb{P}\left(\frac{\mathbf{Z}}{b(t)}\in B_1\cap B_2\right)\longleftrightarrow 0, \quad t\to\infty.$$

2) Suppose $i \in A_1$, $j \in A_2$ and c > 0. Then

(3.7)
$$t\mathbb{P} \quad \frac{Z^{(i)}}{b(t)} > c, \frac{Z^{(j)}}{b(t)} > c \Biggr) \Biggl(\to 0, \quad t \to \infty.$$

3) Angular measure S is concentrated to faces corresponding to A_1 and A_2 ,

(3.8)
$$\mathbb{S}(C^N(A_1)) + \mathbb{S}(C^N(A_2)) = 1$$

Proof. (3.1) \Leftrightarrow 1: Suppose sets $B_1 \subset \mathbb{R}^{N_1}_+$ and $B_2 \subset \mathbb{R}^{N_2}_+$ are bounded away from **0**. Define sets $D_{k,c} \subset \mathbb{R}^N_+$, where $k = 1, 2, \ldots, N$ and c > 0 by

(3.9)
$$D_{k,c} = D_{k,c}^{(1)} \times D_{k,c}^{(2)} \times \ldots \times D_{k,c}^{(N)},$$

where

$$D_{k,c}^{(i)} := \begin{cases} [c,\infty), & i=k\\ \mathbb{R}_+, & i\neq k. \end{cases}$$

Since the sets B_1 and B_2 are bounded away from **0**, there must be numbers $c_1 > 0$ and $c_2 > 0$ so that

$$t\mathbb{P}\left(\frac{\mathbf{Z}_{A_1}}{b(t)} \in B_1, \frac{\mathbf{Z}_{A_2}}{b(t)} \in B_2\right) \leq t\mathbb{P}\left(\underbrace{\mathbf{Z}}_{b(t)} \in (\cup_{k \in A_1} D_{k,c_1}) \cap (\cup_{k \in A_2} D_{k,c_2})\right) \left((3.10) \leq \sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} \left(\mathbb{P}\left(\frac{\mathbf{Z}}{b(t)} \in D_{k_1,c_1} \cap D_{k_2,c_2}\right)\right)\right)$$

Each term on the right hand side of (3.10) converges to 0, as $t \to \infty$ by Condition 1. This shows $1 \Rightarrow (3.1)$. The remaining direction is clear because product sets are special cases of sets in (3.1).

 $1 \Leftrightarrow 2$: Suppose 2 holds. Since B_1 and B_2 are bounded away from **0** there must be indices $k_1 \in A_1$, $k_2 \in A_2$ and a number c > 0 such that $B_1 \subset D_{k_1,c}$ and $B_2 \subset D_{k_2,c}$, where the sets $D_{k_1,c}$ and $D_{k_2,c}$ are defined as in (3.9). Then

$$t\mathbb{P}\left(\frac{\mathbf{Z}}{b(t)}\in B_1\cap B_2\right)\leq t\mathbb{P}\left(\frac{\mathbf{Z}}{b(t)}\in D_{k_1,c}\cap D_{k_2,c}\right)\left(\mathbf{Z}_{k_1,c}\cap D_{k_2,c}\right)$$

where the right hand side converges to 0, as $t \to \infty$ by Condition 2. The remaining direction is clear because the sets in 2 are special cases of sets in 1.

 $3 \Rightarrow 2$: Suppose first that Condition 2 does not hold. Then there exist indices $k_1 \in A_1$, $k_2 \in A_2$ and c > 0 such that (3.7) does not hold, i.e. the limit does not exist or the limit exists but is not 0. Even if the set in (3.7) is a not a continuity set of the limit measure ν , we may choose a smaller number $c' \in (0, c)$ so that the right hand side of

$$\{Z^{(k_1)} > cb(t), Z^{(k_2)} > cb(t)\} \subset \{Z^{(k_1)} > c'b(t), Z^{(k_2)} > c'b(t)\}$$

is a continuity set. So, when c is replaced by c' in (3.7) the limit given by limit measure ν exists, as $t \to \infty$. Since the limit is not 0 by assumption, it must be positive. So, $\nu(D_{k_1,c} \cap D_{k_2,c}) > 0$, where the sets $D_{k_1,c}$ and $D_{k_2,c}$ are as in (3.9). Because the set $D_{k_1,c} \cap D_{k_2,c}$ gets positive value under measure ν , the image under h_1 of this set must have positive angular measure, where h_1 is as in Definition 3.3. Specifically,

(3.11)
$$\mathbb{S}(h_1(D_{k_1,c} \cap D_{k_2,c})) > 0$$

If $x \in h_1(D_{k_1,c} \cap D_{k_2,c})$, then $\sum_{i \in A_1} x^{(i)} > 0$ and $\sum_{i \notin A_2} x^{(i)} > 0$. So,

(3.12)
$$h_1(D_{k_1,c} \cap D_{k_2,c}) \cap C^N(A_1) = \emptyset$$

and

(3.13)
$$h_1(D_{k_1,c} \cap D_{k_2,c}) \cap C^N(A_2) = \emptyset.$$

Since S is a probability measure and some of the probability mass is concentrated outside of the faces $C^{N}(A_{1})$ and $C^{N}(A_{2})$ by (3.11), (3.12) and(3.13), we have that

$$\mathbb{S}(C^{N}(A_{1})) + \mathbb{S}(C^{N}(A_{2})) < 1.$$

So, Condition 3 does not hold.

 $(3.1) \Rightarrow 3$: Suppose Condition 3 does not hold. Then there exist a set $B \subset C^N_+$ such that $\mathbb{S}(B) > 0$,

$$B \cap C^N(A_1) = \emptyset$$

and

$$B \cap C^N(A_2) = \emptyset.$$

Since B does not intersect either simplex, there are numbers $c_1, c_2 \in (0, 1)$ so that the set

$$B_{c_1,c_2} := \left\{ \star \in B : \sum_{i \in A_1} t^{(i)} > c_1, \sum_{i \in A_2} x^{(i)} > c_2 \right\} \left(t^{(i)} > t^{(i)} > t^{(i)} > t^{(i)} \right)$$

has positive angular measure, that is $\mathbb{S}(B_{c_1,c_2}) > 0$. Define $D \subset \mathbb{R}^N$ using B_{c_1,c_2} by $D := \{c\mathbf{x} : c \ge 1, \mathbf{x} \in B_{c_1,c_2}\}$. Now $\nu(D) > 0$. Furthermore,

(3.14)
$$D \subset \left\{ \star \in \mathbb{R}^{N} : \sum_{i \in A_{1}} \star^{(i)} > c_{1}, \sum_{i \in A_{2}} x^{(i)} > c_{2} \right\} \left(19$$

So, when $B_1 = \{ \mathbf{x} \in \mathbb{R}^{N_1} : \sum_i x^{(i)} > c_1 \text{ and } B_2 = \{ \mathbf{x} \in \mathbb{R}^{N_2} : \sum_i x^{(i)} > c_2 \text{ in (3.1) we have that } \}$

$$t\mathbb{P}\left(\frac{\mathbf{Z}_{A_1}}{b(t)} \in B_1, \frac{\mathbf{Z}_{A_2}}{b(t)} \in B_2\right) \ge t\mathbb{P}\left(\frac{\mathbf{Z}}{b(t)} \in D\right)\left($$

where the right hand side does not converge to 0, but to $\nu(D) > 0$. This shows that (3.1) does not hold.

Remark 3.2. Part 2 of Theorem 3.1 admits sets that have zeros in some of their components. For example, if N = 3, $A_1 = \{1, 3\}$ and $A_2 = \{2\}$, then B_1 can be $\{0\} \times \mathbb{R} \times [1, \infty)$. This is why the condition $t\mathbb{P}\left(\mathbf{Z}/b(t) \in [c, \infty)^N\right) \leftrightarrow 0$, as $t \to \infty$ for all c > 0 is not equivalent with asymptotic independence

The following result can be used to reduce multidimensional dependence structures into two dimensional setting by considering sums of components.

Proposition 3.1. Suppose $\mathbf{Z} = (Z^{(1)}, Z^{(2)}, \dots, Z^{(N)})$ is a non negative MRV random vector and $N \geq 2$. Let $A_1, A_2 \subset \{1, 2, \dots, N\}, A_1 \cap A_2 = \emptyset$ and suppose $\#A_1 = N_1$ and $\#A_2 = N_2$, where $N_1, N_2 \ge 1$ and $N_1 + N_2 = N$. Then the non negative two dimensional random vector

$$(Y_1, Y_2) := \left(\sum_{i \in A_1} Z^{(i)}, \sum_{i \in A_2} Z^{(i)} \right) \left(\left(\sum_{i \in A_1} Z^{(i)} \right) \right) \right)$$

is also MRV. Furthermore, \mathbf{Z}_{A_1} and \mathbf{Z}_{A_2} are asymptotically independent if and only if Y_1 and Y_2 are asymptotically independent.

Proof. The fact that (Y_1, Y_2) is MRV follows from Proposition 5.5 found from [22, p. 142].

For the latter claim, observe first using (3.3) that if $\mathbf{x} \in C^N_+$, then for $j = 1, 2, \boldsymbol{Q}_{\mathbf{a}_1, \mathbf{a}_2}(\mathbf{x}) = \mathbf{a}_j$ if and only if $\mathbf{x} \in C(A_j)$. So, it follows that $S(C(A_1)) + S(C(A_1)) = 1$ if and only if $S_Y((0,1)) + S_Y((1,0)) = 1$, where \mathbb{S}_{Y} denotes the angular measure of (Y_1, Y_2) . Using Part 3 of Theorem 3.1 completes the proof.

Asymptotic normality of the validation statistic 3.3

We start by defining an auxiliary function g function in Definition 3.4 for Theorem 3.2. Function q is used to fix a set on simplex C^2 . It is then tested if the asymptotic support is included in the fixed set. Different choices for

g yield tests for different dependence structures. These structures include asymptotic independence introduced in Section 3, but the construct allows other choices as well. The most commonly encountered structures are illustrated in Figure 1.

Definition 3.4. Suppose $[a_1, b_1], [a_2, b_2], \ldots, [a_m, b_m]$ are separate subintervals of [0, 1], where $m \ge 2$.

Let $g: [0,1] \to \mathbb{R}$ be a function defined by conditions

$$g(0) = \begin{cases} q(i) & a_1 > 0\\ \frac{1}{2}, & a_1 = 0, \end{cases}$$
$$g(a_i) = g(b_i) = \frac{1}{2} + \frac{i-1}{2(m-1)}, i = 1, 2, \dots, m,$$
$$g((b_i + a_{i+1})/2) = g(b_i) - \frac{1}{2}, i = 1, 2, \dots, m-1$$

and

$$g(1) = \begin{cases} \frac{1}{2}, & b_m < 1\\ 1, & b_m = 1 \end{cases}$$

and whose values are given by linear interpolation between the defined points on the rest of the interval [0, 1].

The function g enables the user to add small buffers in which the support must lie. The feature is added because, to our experience, it is challenging to detect asymptotic independence from real data. The task can be made easier if one admits small deviation from true asymptotic independence by widening the search for support measures S that concentrate near the axes but not necessarily on the axes themselves. Such support structures can still convey useful information. This is because they imply that some of the components can not obtain large values at the same time which is precicely the needed information in many applications. Similar approaches for finding sufficiently independent groups of variables exist in the literature, for example in [10].

In practice, the most frequently searched dependence structures correspond to asymptotic independence and strong asymptotic dependence. Tests for these are presented in Remarks 3.3 and 3.4 below. We prove first a more general result from which the others follow. The results are formulated for positive vectors for notational simplicity.



Figure 1: Graphs of function g for different test scenarios. On the left, g corresponds to a setting where asymptotic independence is tested with buffers. Values at end points differ in order to avoid zero variance of L in (3.16) under asymptotic independence. In the middle, g could be used to test if the asymptotic support is covered by two intervals. In addition, similar g could arise when testing if the support is covered by a single interval after the sample is processed using the method described in Remark 3.4. On the right, g tests if the support is covered by three separate intervals. Such dependence structure might arise e.g. in the search of hidden regular variation after the first order cone is removed from data.

Theorem 3.2 (Asymptotic normality of test statistic). Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be i.i.d MRV random vectors in \mathbb{R}^2_+ . Suppose $(R_i, \theta_i) \in \mathbb{R}_+ \times C^2_+$ is the polar coordinate representation of \mathbf{Z}_i , where $R_i = ||\mathbf{Z}_i||$ and $\theta_i = \mathbf{Z}_i/||\mathbf{Z}_i||$. Let $\theta_{(i:n)} = (\theta^1_{(i:n)}, \theta^2_{(i:n)}) = (\theta^1_{(i:n)}, 1 - \theta^1_{(i:n)})$ be the angular component of the ith largest vector in L_1 norm out of a sample whose size is n. Suppose $m \geq 2$ and g is as in Definition 3.4. Let \mathbb{S}_1 be a probability measure on [0,1] obtained as a push forward measure from the angular measure \mathbb{S} via mapping $(x, y) \mapsto x$. Assume $\mathbb{S}_1(\cup_i[a_i, b_i]) = 1$.

Denote

$$\mu_g := \iint_{\mathbb{Q}}^{\mathbb{L}} g(x) \, \mathbb{S}_1(\mathrm{d}x) = \sum_{i=1}^m \left(\frac{1}{2} + \frac{i-1}{2(m-1)} \right) \mathbb{S}_1([a_i, b_i])$$

and

$$\sigma_g^2 := \iint_0^1 (g(x) - \mu_g)^2 \, \mathbb{S}_1(\mathrm{d}x) = \sum_{i=1}^m \left(\frac{1}{2} + \frac{i-1}{2(m-1)} - \mu_g \right)^2 \, \mathbb{S}_1([a_i, b_i]).$$
If

(3.15)
$$\sqrt{k(n)} \left(\mathbb{E}(g(\theta^{1}_{(k(n):n)})) - \mu_{g} \right) \longleftrightarrow 0$$

and $k(n)/n \to 0$, as $n \to \infty$, then

(3.16)
$$\hat{T} := \frac{\sum_{i=1}^{k(n)} \left(g(\theta_{(i:n)}^1) - \mu_g \right)}{(k(n))^{\frac{1}{2}} \sigma_g} \left(\stackrel{d}{\to} N(0, 1), \right.$$

as $n \to \infty$.

Proof. The proof is similar to the proof of Theorem 3 of [21] and follows from it with minor modifications. \Box

Remark 3.3. If m = 2, $a_1 = b_1 = 0$ and $a_2 = b_2 = 1$ in Theorem 3.2, then (3.16) is a test statistic for asymptotic independence.

Remark 3.4. If the limiting measure S is concentrated into a single interval $[a, b] \subsetneq [0, 1]$, Theorem 3.2 can not be directly applied because it requires the limiting variable L to have a non zero variance. However, the case where the asymptotic support is an interval can be reduced to the setting of two intervals by first transforming the sample $(Z_i^{(1)}, Z_i^{(2)})_{i=1}^n$. We can assume the sample size n is even. If it is not, we can leave out

We can assume the sample size n is even. If it is not, we can leave out the observation with the smallest L_1 norm, because it has no effect to the subsequent analysis. When i is odd, transform the two dimensional data using mapping $(x, y) \mapsto (x/2, x/2 + y)$. If i is even, use mapping $(x, y) \mapsto$ (x + y/2, y/2) instead. Then permute the order of observations to obtain i.i.d MRV random vectors. The limiting measure of the transformed sample replaces the original with two smaller copies. In addition, $\operatorname{supp}^+ \subset [a, b]$ if and only if the asymptotic support of the transformed sample is covered by $[a/2, b/2] \cup [(a + 1)/2, (b + 1)/2].$

Remark 3.5. Since it is assumed in Proposition 3.2 that all probability mass of S is concentrated into intervals $[a_1, b_1], [a_2, b_2], \ldots, [a_m, b_m]$ and g is constant on those intervals, the random variable L can obtain at most m different values. In practice probabilities $\mathbb{P}(L = g(a_i)), i = 1, 2, \ldots, m$ need to be estimated.

3.3.1 Discussion on the choice of g in Definition 3.4

The purpose of Function g in Definition 3.4 is to identify when the sets $[a_1, b_1], [a_2, b_2], \ldots, [a_m, b_m]$, called the *test intervals*, eventually cover the support **supp**⁺, as $n \to \infty$. Selecting a function with best performance in terms of a pre set benchmark depends from the way and rate at which convergence to the limit measure takes place. In practical scenarios such

information is not usually available. Our suggestion for g is chosen by empirical testing with different data sets.

While there are multiple ways to define such functions, the construction of a limiting result corresponding to Proposition 3.2 sets some requirements that restrict the set of possible choices. The rationale for choosing g as it is defined in Definition 3.4 is the following. Function g must be set to a constant value on all separate intervals that are believed to contain probability mass of S. One might be tempted to use a symmetric function around 1/2 so that the values in both endpoints of [0, 1] would be the same. The problem with this is that it would also imply zero variance for a variable analogous to L in Theorem 3.2 in the case of asymptotic independence, making it unclear how the test quantity should be scaled for a non-degenerate limiting distribution. Once it is established that the function cannot be symmetric, the different values in endpoints are not significant since the system can be scaled and shifted to the values set in Definition 3.4.

The remaining question is then how the function g should behave between the regions of constant value. Firstly, Function g should be able to separate desirable distributions from the ones with support that is not concentrated on the test intervals. A way to do this is to make the quantity $|\hat{T}|$ of (3.16) as large as possible in the presence of unwanted limiting behavior. On the other hand, the thresholded data may contain pre-limit observations whose projections are not in \mathbf{supp}^+ even when all limiting probability mass on simplex is covered by the test intervals. So, observations close to the regions of constant value should not change the value of $|\hat{T}|$ too dramatically. In conclusion, the choice of g in Definition 3.4 seems to be a reasonable compromise between the two opposing goals. It is chosen from the class of piece-wise linear functions for computational simplicity.

4 Examples with simulated and real data

In this section, we illustrate how the theoretical results concerning support estimation in Section 2 and support testing in Section 3 can be used in practice. We begin with a simulated dataset in Example 4.1 to show how the grid based support estimator performs in a controlled environment. Example 4.2 studies daily stock returns. The emphasis is on the fact that stocks in the same field tend to be dependent, but one can find at least asymptotically independent assests among ordinarily listed equities. In Example 4.3, a natural scenario for emergence of asymptotic independence is given using rainfall data¹. Finally, in Example 4.4 daily returns of gold and silver are used to show how the support estimates can be used to obtain inequalities for sizes of large fluctuations.

Typically, multivariate datasets require some amount of processing before they can reasonably be thought to satisfy assumptions of multivariate regular variation given in Definition 1.1. In particular, tail indices of marginal distrtibutions must to be the same for the asymptotic theory to work. To this end, one needs to estimate tail indices. Estimation of tail index is a classical topic which is discussed e.g. in [1,22,25] or more recently in [13]. If it is decided that the marginals do not have the same index, then the data needs to be transformed before proceeding further.

Multiple methods exist for transforming datasets to fit the scope of multivariate regular variation. Usual methods include power transformations of marginals or the rank transform, see [22, Section 9.2] and [11].

4.1 Simulated data

Support estimator of Section 2 is applied to simulated data. The data set consists of 3 dimensional observations $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$, where n = 150000. Observations are generated by fixing a region $A \subset C_+^3$ and then sampling uniformly 50000 samples from A. The samples on the simplex are then assigned a radial component independently from Pareto(2) distribution. So, by definition, the angular and radial components of the observations are independent. Additionally, 100000 observations are added to the sample depicting noise by sampling uniformly from the entire simplex C_+^3 and assigning them with an exponentially distributed radial components. Finally, we put the simulated samples into a random order so that they form an i.i.d sample from a mixture distribution that is MRV.

Figure 2 Illustrates how well the grid based support estimate is able to find the location of the set A. The dots in figures 2a and 2c are projected k = 10000 largest observations in L_1 . The dark dense region is the set A, which is a circle in 2a and a triangle in 2c. In figures 2b and 2d the set A is estimated by forming the support estimator $A_{k,m,q}$ using parameter values k = 10000, m = 36 and q = 0.01. Rejecting some of the points by positive q produces clearly visible rasterised version of A with no misidentified cells. This is due to the fact that our simulated data fits perfectly to the MRV framework. The following examples show that real data produces far less conclusive results.

¹Special thanks are due to Sebastian Engelke who suggested that asymptotic independence could be found from rainfall data in a personal communication.



Figure 2: Figures present projected and estimated supports of simulated data. The large red triangles indicate the boundaries of the image of C_{+}^{3} under a simplex mapping T discussed in Section 1.4.

4.2 Stock data vs. catastrophe fund

Stock market dependencies are studied using a data set consisting of daily prices of 6 stocks and a catastrophe fund. The studied equities and their ticker symbol abbreviations are: Google (GOOG), Microsoft (MSFT), Apple (AAPL), Chevron (CVX), Exxon (XOM), British Petrol (BP) and CATCo Reinsurance Opportunities Fund (CAT.L). Observations range from December 20, 2010 to 10 July, 2018. The data set was downloaded via R package *Quantmod*.

Observations were processed by the taking logarithm and calculating differences. The resulting components of the data set have similar tail indices with positive and negative tails. However, the index of CAT.L was substantially smaller than the others, making it necessary to use rank transform when comparing it against the other equities.

In Figure 3, the strength of pairwise dependence is calculated using the largest k = 200 observations in L_1 projected to C_+^2 , denoted $z_1, z_2, \ldots, z_{200}$, by formula $\sum_{i=1}^{k} (1 - d_2(1/2, z_i)/k)$. That is, the observations that have the largest distance to the midpoint (1/2, 1/2) of simplex C_+^2 are assigned small values and observations near the midpoint large values. All pairwise dependencies that exceed the level 0.46 are drawn as edges in Figure 3. The level was obtained empirically by gradually lowering the required level and observing which connections appeared on the graph first, that is, which dependencies are the strongest.

Figure 3 suggests that companies within the same financial sector, oil or technology, might not be asymptotically asymptotically independent. However, the catastrophe fund might be asymptotically independent from stocks.

In Figure 4, projected and estimated supports of positive orthants are



Figure 3: A graph illustrating the strongest preliminary pairwise asymptotic dependencies.

depicted for oil and tech stocks. Parameter values are k = 200, m = 12 and q = 0.01. Estimated grid based supports in subfigures 4b and 4d suggest that the groups of stocks are not asymptotically independent, but possibly quite dependent. However, based on Figure 3 oil and tech sectors might be asymptotically independent and CAT.L could be asymptotically independent of all the studied stocks.

Asymptotic independence was tested using absolute values of observations. Function g was defined using choices $m = 2, a_1 = 0, b_1 = 0.1, a_2 = 0.9$ and $b_2 = 1$, i.e. asymptotic independence was tested with buffers. It was assumed that $S_1([0, 0.1]) = S_1([0.9, 1]) = 1/2$. Observations from oil and tech fields were added together in order to form a two dimensional vectors. The empirical test statistic \hat{t} corresponding to \hat{T} of (3.16) was calculated: $\hat{t} \approx 0.73$. So, under null hypothesis $\mathbb{P}(\hat{T} > |\hat{t}|) \approx 0.31$ and thus the test statistic is consistent with the idea that oil and tech sectors could be asymptotically independent. Similar test was performed pairwise with CAT.L against all 6 stocks. There was not enough evidence based on the test statistics corresponding to (3.16) to reject the idea of asymptotic independence.

In conclusion, the test statistic developed in Section 3 supports what is seen in the preliminary dependence graph, in Figure 3.



Figure 4: Projected largest k = 200 observations and the estimated supports for tech stocks, in Figures 4a-4b, and for oil stocks, in Figures 4c-4d, respectively.



Figure 5: Projected and estimated supports of tech stocks.

4.3 FMI data

Daily rainfall data was downloaded from Finnish Meteorological Institute from 3 separate locations. Only observations from summer months June, July and August were taken into consideration to reduce seasonal effects. Two of the locations, Kouvola and Savonlinna were close to each other where as the last one, Sodankyla, was further away. Rainfalls in the nearby locations showed high dependence. The rainfalls of the further location were not entirely independent of the two others, but exhibited independence in the largest observations. The total number of observations in the data is n = 3864.

In figure 6, projected and estimated supports of rank transformed rainfall vectors are presented using k = 300 larges observations. In the support estimate, parameter values m = 12 and q = 0.01 were used. The rainfall data seems to support the idea that locations in close proximity are fully dependent and locations far away from each other are independent. Between the two extreme cases when the distance of the studied locations is suitable,



Figure 6: Projected and estimated supports of daily rainfall data recorded in 3 locations in Finland.

rainfalls become asymptotically independent.

4.4 Gold vs Silver price data

In this section we analyze daily gold and silver price data. The data is gathered from London Bullion Market Association and it is downloaded via R package *Quandl*. In the data, the price of one ounce of gold or silver is recorded each day during a time period ranging from December 3, 1973 to January 15, 2014. Only complete cases where the price information was available from both gold and silver were accepted as part of the data set. There were three days where price information was incomplete. Large price fluctuations did not occur during the omitted days and thus ignoring them has no effect to the resulting asymptotic analysis.

The daily price data was transformed by logarithmic differentiation in order to obtain a sample which is better suited with the i.i.d. assumption of the model. The individual positive and negative marginals of gold and silver seemed to be reasonably in line with the assumption of regular variation. No power or rank transformations were performed to the data set. The resulting sample of n = 10323 was thresholded by the k = 200 largest observations in L_1 norm and then projected onto C^2 to produce the diamond plot presented in Figure 7.



Figure 7: Diamond plot of daily price data of gold and silver after logdiff transformation. Horizontal axis corresponds to gold and vertical axis to silver.

Figure 7 shows that the largest fluctuations in gold and silver prices tend to occur to the same direction. In addition, it seems that the points do not fill the positive or negative quadrant of the C^2 simplex evenly, but concentrate on intervals. The estimation of asymptotic support in the negative quadrant was chosen as a suitable example, analysis of other quadrants could be performed similarly. So, only the part of data where both components are negative was used. The n = 3951 observations were multiplied by -1 to obtain a data set int the positive quadrant.

The one dimensional grid based estimator was obtained using the first 1975 observations sampled uniformly without replacement from the data. The points were projected using a simplex mapping $T: C_+^2 \to [0, 1]$ defined by T(x, y) = x. Since gold is on the horizontal axis, the projected values on [0, 1] close to 0 correspond to silver and values near 1 to gold. The grid based support estimator with parameter values n = 1975, k = 100, m = 15 and q = 0.02 suggests that the asymptotic support should be covered by

interval [0, 0.65].



Figure 8: In Figure 8a, the data set transformed using method of Remark 3.4. The transformed observations are presented in Figure 8c. Figures 8b and 8d show the diamond plot of the k = 100 largest observations in L_1 norm.

The validity of the support estimate was tested using the remaining 1976 observations. Function g was formed using the method discussed in Remark 3.4. The process is illustrated in Figure 8. The aim is to test if the asymptotic support of the transformed data is covered by $[0, 0.325] \cup [0.5, 0.825]$. The null hypothesis is that in our sample $\hat{T} \sim N(0, 1)$ where \hat{T} is as in Equation (3.16). The empirical test statistic \hat{t} corresponding to \hat{T} quantity was calculated from the remaining observations with result $\hat{t} \approx 0.074$. Under null hypothesis $\mathbb{P}(\hat{T} > |\hat{t}|) \approx 0.398$. So, the value of \hat{t} gives no reason to think the null hypothesis is wrong and the asymptotic support of the original sample could well be covered by the set [0, 0.65].

As a practical application we immediately obtain inequalities for large fluctuations in gold and silver prices. Denote the daily logarithmic decrease in prices with x for gold and y for silver. If a very large decrease is observed for gold, i.e. x is large, then the support estimate implies $x/(x+y) \leq 0.65$ so that $y \geq 0.53x$. In other words, the support estimate says it is unlikely for the decrease in logarithmic silver price to be less than 0.53x. The method allows estimation of quantities that are unknown in the presence of extremal circumstances.

4.5 Final thoughts

Based on the previous examples, the methods presented in Sections 2 and 3 seem to be usable in some scenarios. However, the use of asymptotic support estimation has its limitations. For one, it is challenging to find a large sample of vectors with tail equivalent marginals that satisfies the i.i.d.

assumption. With time series, larger number of observations may lead to poor results because the underlying mechanisms that produce observations may change when the data is gathered during a long time. So, the initial data needs to be normalized to fit the theoretical framework. In financial context, one popular method is *de-GARCHing*, see [12, Sec 2.1.]. The choice of pre processing method adds a new source of uncertainty to the model.

Additional problems arise when the marginals are not tail equivalent and especially when the rank transform is used. If the original sample has MRV distribution, it is known that the rank transformed sample should produce the same asymptotic support as the original sample in the limit as the number of observations grows. However, it is not known to the authors how the support estimate based on a finite rank transformed sample differs from the true asymptotic support. This raises the question about the sufficient number of observations needed for the asymptotical analysis to be reasonable. A related problem is discussed in [18].

The proposed method in Sections 2 and 3 has an exploratory component involved, as the support estimate is based on parameters whose values need to be chosen. In practice usual plotting methods turned out to be inefficient in the choice of parameters. This is why a dedicated software is being developed for the purpose of asymptotic support identification and testing.

Acknowledgements

Jaakko Lehtomaa gratefully acknowledges hospitality and use of the services and facilities of ORIE School at the Cornell University during his visit from Sep 2017 to Aug 2018, funded by the Finnish Cultural Foundation via the Foundations' Post Doc Pool. Sidney Resnick was partially supported by US ARO MURI grant W911NF-12-1-0385 to Cornell University.

References

- CLAUSET, A., SHALIZI, C. R., AND NEWMAN, M. E. J. Power-law distributions in empirical data. SIAM Rev. 51, 4 (2009), 661–703.
- [2] DAS, B., MITRA, A., AND RESNICK, S. Living on the multidimensional edge: seeking hidden risks using regular variation. Adv. in Appl. Probab. 45, 1 (2013), 139–163.
- [3] DAS, B., AND RESNICK, S. I. Models with hidden regular variation: generation and detection. Stoch. Syst. 5, 2 (2015), 195–238.
- [4] DAS, B., AND RESNICK, S. I. Hidden regular variation under full and strong asymptotic dependence. *Extremes 20*, 4 (2017), 873–904.
- [5] EINMAHL, J. H. J., DE HAAN, L., AND SINHA, A. K. Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* 70, 2 (1997), 143–171.

- [6] EINMAHL, J. H. J., AND SEGERS, J. Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. Ann. Statist. 37, 5B (2009), 2953–2989.
- [7] EMBRECHTS, P., HOFERT, M., AND WANG, R. Bernoulli and tail-dependence compatibility. Ann. Appl. Probab. 26, 3 (2016), 1636–1658.
- [8] EMBRECHTS, P., KLÜPPELBERG, C., AND MIKOSCH, T. Modelling extremal events, vol. 33 of Applications of Mathematics (New York). Springer-Verlag, Berlin, 1997. For insurance and finance.
- [9] FRAHM, G., JUNKER, M., AND SCHMIDT, R. Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance Math. Econom.* 37, 1 (2005), 80–100.
- [10] GOIX, N., SABOURIN, A., AND CLÉMENÇON, S. Sparse representation of multivariate extremes with applications to anomaly detection. J. Multivariate Anal. 161 (2017), 12–31.
- [11] HEFFERNAN, J., AND RESNICK, S. Hidden regular variation and the rank transform. Adv. in Appl. Probab. 37, 2 (2005), 393–414.
- [12] HOFERT, M., AND OLDFORD, W. Visualizing dependence in high-dimensional data: An application to S&P 500 constituent data. *Econometrics and Statistics* (2017).
- [13] KIM, J. H., AND KIM, J. A parametric alternative to the hill estimator for heavytailed distributions. *Journal of Banking & Finance* 54 (2015), 60 – 71.
- [14] LARSSON, M., AND RESNICK, S. I. Extremal dependence measure and extremogram: the regularly varying case. *Extremes* 15, 2 (2012), 231–256.
- [15] LEHTOMAA, J. Limiting behaviour of constrained sums of two variables and the principle of a single big jump. *Statist. Probab. Lett.* 107 (2015), 157–163.
- [16] LINDSKOG, F., RESNICK, S. I., AND ROY, J. Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. Probab. Surv. 11 (2014), 270–314.
- [17] MATSUI, M., AND MIKOSCH, T. The extremogram and the cross-extremogram for a bivariate GARCH(1,1) process. Adv. in Appl. Probab. 48, A (2016), 217–233.
- [18] NGUYEN, T., AND SAMORODNITSKY, G. Tail inference: where does the tail begin? Extremes 15, 4 (2012), 437–461.
- [19] PENG, L. Estimation of the coefficient of tail dependence in bivariate extremes. Statist. Probab. Lett. 43, 4 (1999), 399–409.
- [20] RESNICK, S. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* 5, 4 (2002), 303–336 (2003).
- [21] RESNICK, S. The extremal dependence measure and asymptotic independence. Stoch. Models 20, 2 (2004), 205–227.
- [22] RESNICK, S. I. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. Probabilistic and statistical modeling.
- [23] RUSSELL, B. T., COOLEY, D. S., PORTER, W. C., REICH, B. J., AND HEALD, C. L. Data mining to investigate the meteorological drivers for extreme ground level ozone events. Ann. Appl. Stat. 10, 3 (2016), 1673–1698.
- [24] SCHÖLKOPF, B., PLATT, J. C., SHAWE-TAYLOR, J. C., SMOLA, A. J., AND WILLIAMSON, R. C. Estimating the support of a high-dimensional distribution. *Neural Comput.* 13, 7 (July 2001), 1443–1471.

[25] VIRKAR, Y., AND CLAUSET, A. Power-law distributions in binned empirical data. Ann. Appl. Stat. 8, 1 (2014), 89–119.