



HAL
open science

Dependence Testing via Extremes for Regularly Varying Models

Malcolm Egan

► **To cite this version:**

Malcolm Egan. Dependence Testing via Extremes for Regularly Varying Models. ISIT 2021 - IEEE International Symposium on Information Theory, Jul 2021, Melbourne, Australia. pp.1-6, 10.1109/ISIT45174.2021.9518201 . hal-03223444

HAL Id: hal-03223444

<https://hal.science/hal-03223444>

Submitted on 11 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dependence Testing via Extremes for Regularly Varying Models

Malcolm Egan

Abstract—In heavy-tailed data, such as data drawn from regularly varying models, extreme values can occur relatively often. As a consequence, in the context of hypothesis testing, extreme values can provide valuable information in identifying dependence between two data sets. In this paper, the error exponent of a dependence test is studied when only processed data recording whether or not the value of the data exceeds a given value is available. An asymptotic approximation of the error exponent is obtained, establishing a link with the upper tail dependence, which is a key quantity in extreme value theory. While the upper tail dependence has been well characterized for elliptically distributed models, much less is known in the non-elliptical setting. To this end, a family of non-elliptical distributions with regularly varying tails arising from shot noise is studied, and an analytical expression for the upper tail dependence derived.

I. INTRODUCTION

Testing for dependence between two data sets is a fundamental problem in statistical inference [1], [2], arising in biology to study evidence of associations between different molecular species (see e.g., [3]) and also in channel coding [4]. The simplest class of dependence testing problems arises in the form of a binary hypothesis test [5], which aims to distinguish between independent data and data drawn from a given joint distribution.

One variation on this basic setting is when the data is processed as follows. Suppose that the data is the sequence of independent and identically distributed pairs $\{(X_k, Y_k)\}_{k=1}^n$ on \mathbb{R}^2 and that the processed data is defined by $U_k = \mathbf{1}_{\{X_k > \nu\}}$ and $V_k = \mathbf{1}_{\{Y_k > \nu\}}$ for a given threshold $\nu > 0$. That is, U_k, V_k are binary variables that take the value one if the data X_k, Y_k , respectively, are greater than the threshold ν and zero otherwise. In this case, dependence can only be detected from extreme values of the data, which typically reduces required compression rates in the context of decentralized tests [6].

A dependence test can then be defined for the processed data. By the Chernoff-Stein lemma, the corresponding error exponent is given in terms of the Kullback-Leibler divergence between the joint and product distributions of the processed data [7], when each sample is independent. A simple observation is that if ν is very large, the error exponent will be close to zero. Nevertheless, the rate at which the error exponent tends to zero is sensitive to the tail behavior of the data (X_k, Y_k) . In particular, it is natural to expect that when (X_k, Y_k) are heavy tailed, the error exponent will be larger than for lighter tailed distributions for a fixed threshold ν .

In this paper, we study the behavior of the error exponent for this dependence test based only on extreme values of the data $\{(X_k, Y_k)\}$; i.e., as the threshold $\nu \rightarrow \infty$. This perspective differs from existing work, which does not typically consider the impact of data pre-processing due to, for example, communication constraints (e.g., [1], [2]). We focus on the case that the data has regularly varying tails [8], which, roughly speaking, means that the probability density function of the data decays polynomially. Our first observation is that the rate the error exponent tends to zero depends on the upper tail dependence of the data, which characterizes the probability Y_k is large conditioned on the event that X_k is also large. While the upper tail dependence plays a key role in extreme value theory [9], it appears to have not previously arisen in the evaluation of an information measure. As a byproduct, we also obtain a new characterization of the upper tail dependence in terms of a Kullback-Leibler divergence.

The upper tail dependence has a well-known characterization for elliptical distributions with regularly varying marginals. As such, our characterization of the error exponent in the dependence test based on extremes can be readily evaluated for this family of distributions. However, there are far fewer characterizations for the upper tail dependence for non-elliptical distributions. To this end, we study a family of non-elliptical distributions with regularly varying marginals induced by shot noise, which is closely related to interference statistics in wireless communications [10]. For this family, we derive a new analytical expression for the upper tail dependence.

A. Notation

Vectors are denoted by bold lowercase letters and random vectors by bold uppercase letters, respectively (e.g., \mathbf{x}, \mathbf{X}). We denote the distribution of a random vector \mathbf{X} by $P_{\mathbf{X}}$. The cumulative distribution function of random variable X is denoted by $F_X(\cdot)$, and the complementary cumulative distribution function by $\bar{F}_X(\cdot) = 1 - F_X(\cdot)$. Let \mathbf{X} and \mathbf{Y} be continuous random vectors on \mathbb{R}^d with distributions $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$, respectively. The density of \mathbf{X} is denoted by $p_{\mathbf{X}}$ (w.r.t the Lebesgue measure) and the density of \mathbf{Y} (w.r.t the Lebesgue measure) by $p_{\mathbf{Y}}$, when they exist. If \mathbf{X}, \mathbf{Y} are two random vectors equal in distribution, then we write $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$. The Kullback-Leibler divergence between two distributions $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$ on \mathbb{R}^d is then defined by

$$D(P_{\mathbf{X}}||P_{\mathbf{Y}}) = \int_{\mathbb{R}^d} p_{\mathbf{X}}(\mathbf{x}) \log \frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{x})} d\mathbf{x}. \quad (1)$$

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$. We use the Landau notation where $f(x) = o(g(x))$ if $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$. For two functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$, we write $f(x) \sim g(x)$, $x \rightarrow \infty$ if $f(x) = g(x) + o(g(x))$, $x \rightarrow \infty$.

II. PROBLEM FORMULATION

Consider a sequence of continuous n independent and identically distributed random vectors $\{(X_k, Y_k)\}_{k=1}^n$ on \mathbb{R}^2 and let $\mathcal{H}_0, \mathcal{H}_1$ be the hypotheses that the X_k and Y_k are independent and dependent with joint density p_{XY} , respectively. Suppose that X and Y have the same distribution. In order to test for dependence, by the Neyman-Pearson lemma, the optimal test for a given type I error probability α has the form

$$\prod_{k=1}^n \frac{p_{XY}(X_k, Y_k)}{p_X(X_k)p_Y(Y_k)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{>}} \tau_\alpha, \quad (2)$$

where τ_α is chosen to satisfy the type I error constraint.

Let $\nu > 0$. Another test can then be constructed using the processed data

$$U_k = \mathbf{1}_{\{X_k > \nu\}}, \quad V_k = \mathbf{1}_{\{Y_k > \nu\}}, \quad k = 1, \dots, n. \quad (3)$$

Let $\mathcal{A} \subset \{0, 1\}^n$ be a decision region for \mathcal{H}_1 . For a fixed $0 < \epsilon < \frac{1}{2}$ and type I error probability $\alpha_n < \epsilon$, the minimal type II error is given by

$$\beta_n^\epsilon = \min_{\substack{\mathcal{A} \subset \{0, 1\}^n \\ \alpha_n < \epsilon}} \beta_n. \quad (4)$$

The Chernoff-Stein lemma [7, Theorem 11.8.3] then yields an error exponent

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^\epsilon = -D(P_{UV} || P_U P_V). \quad (5)$$

The problem that we consider in the remainder of this paper is to characterize $D(P_{UV} || P_U P_V)$ with a focus on the case where X and Y have regularly varying distributions. Recall that in the univariate case, a random variable Z has a regularly varying distribution with index $-a \in \mathbb{R}$ and slowly varying function L if

$$\mathbb{P}(|Z| \geq x) \sim L(x)x^{-a}, \quad x \rightarrow \infty, \quad (6)$$

where, for all $t > 0$, $L : \mathbb{R} \rightarrow \mathbb{R}$ satisfies

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1. \quad (7)$$

III. AN ASYMPTOTIC CHARACTERIZATION OF

$$D(P_{UV} || P_U P_V)$$

In this section, we derive a closed-form asymptotic characterization of $D(P_{UV} || P_U P_V)$ for large ν . First, recall that

$$\begin{aligned} & D(P_{UV} || P_U P_V) \\ &= \sum_{m=0}^1 \sum_{k=0}^1 \mathbb{P}(U = k, V = m) \log \frac{\mathbb{P}(V = m | U = k)}{\mathbb{P}(V = m)}. \end{aligned} \quad (8)$$

A quantity that will play a central role in our analysis is the upper tail dependence of (X, Y) , which is given by (see e.g., [11])

$$\begin{aligned} \lambda &= \lim_{\nu \rightarrow \infty} \mathbb{P}(Y > \nu | X > \nu) \\ &= \lim_{\nu \rightarrow \infty} \mathbb{P}(V = 1 | U = 1). \end{aligned} \quad (9)$$

The upper tail dependence can also be expressed in terms of the copula function associated with (X, Y) . Recall that for a continuous random vector (Z_1, Z_2) , the joint distribution can be written as

$$\Pr(Z_1 \leq z_1, Z_2 \leq z_2) = C(F_1(z_1), F_2(z_2)), \quad (10)$$

where $C : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a copula function, and F_1, F_2 are the marginal distributions of Z_1, Z_2 , respectively. By Sklar's theorem [11], for each distribution $\Pr(Z_1 \leq z_1, Z_2 \leq z_2)$, there exists a unique copula function $C(\cdot)$. We then have the representation of the upper tail dependence given by [11]

$$\lambda = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}. \quad (11)$$

As such, the upper tail dependence does not depend on the marginal distributions F_1, F_2 and captures the dependence structure of (Z_1, Z_2) .

The upper tail dependence is known for several families of multivariate distributions (including those that are not regularly varying) [12]. For example,

- (i) Multivariate Gaussian distributions (non-regularly varying): $\lambda = 0$.
- (ii) Distributions with a known copula function: λ given in (11).
- (iii) Elliptical distributions with regularly varying marginals:

$$\lambda = \frac{\int_0^{\frac{1}{\sqrt{2}}} \frac{u^a}{\sqrt{1-u^2}} du}{\int_0^1 \frac{u^a}{\sqrt{1-u^2}} du}, \quad (12)$$

where $-a$ is the index of regular variation.

Recall that a random vector (Z_1, Z_2) is said to have an elliptical distribution if its characteristic function has the form

$$\Psi_{Z_1, Z_2}(\boldsymbol{\theta}) = \varphi(\boldsymbol{\theta}^T R \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^2 \quad (13)$$

for some positive definite matrix $R \in \mathbb{R}^{2 \times 2}$ and function $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ such that Ψ_{Z_1, Z_2} is a valid characteristic function. Important examples of elliptical distributions with regularly varying tails are the multivariate t -distribution and the sub-Gaussian α -stable distribution.

For large ν in (3), $D(P_{UV} || P_U P_V)$ admits a concise asymptotic approximation in terms of the upper tail dependence.

Proposition 1. *Suppose that P_Y, P_X are regularly varying with index $-a$ and slowly varying function L . If $\lambda > 0$, then*

$$D(P_{UV} || P_U P_V) \sim \lambda L(\nu) \nu^{-a} \log \nu^a, \quad \nu \rightarrow \infty. \quad (14)$$

Proof. Substitute (6) and (9) in (8) to obtain

$$\begin{aligned}
& D(P_{UV}||P_U P_V) \\
& \sim \lambda L(\nu)\nu^{-a} \log\left(\frac{\lambda\nu^a}{L(\nu)}\right) + (1-\lambda)L(\nu)\nu^{-a} \log(1-\lambda) \\
& \quad + (1-\lambda)L(\nu)\nu^{-a}(1-L(\nu)\nu^{-a}) \log\left(\frac{1-\lambda}{1-L(\nu)\nu^{-a}}\right) \\
& \quad + (1-(1-\lambda)L(\nu)\nu^{-a})(1-L(\nu)\nu^{-a}) \\
& \quad \cdot \log\left(\frac{1-(1-\lambda)L(\nu)\nu^{-a}}{1-L(\nu)\nu^{-a}}\right), \quad \nu \rightarrow \infty \tag{15}
\end{aligned}$$

then take the dominant term. \square

Observe that the asymptotic expression for $D(P_{UV}||P_U P_V)$ in Proposition 1 depends explicitly on the upper tail dependence λ . While the result is straightforward to prove, this is the only example that we are aware of where the upper tail dependence naturally arises in the evaluation of an information theoretic measure. This is despite the importance of the upper tail dependence in characterizing the dependence structure of heavy-tailed random vectors [9].

Note that a refined approximation of $D(P_{UV}||P_U P_V)$ is given by (15). We also point out that when (X_k, Y_k) is a Gaussian random vector, a similar expansion can be developed in terms of the corresponding exponentially decaying tail behavior. However, the dominant term is not the same as in Proposition 1 since the upper tail dependence is zero for multivariate Gaussian random vectors.

To illustrate the behavior of the asymptotic approximation, consider the scenario where (X, Y) is an isotropic sub-Gaussian α -stable random vector ($\alpha < 2$). Recall that a random vector \mathbf{Z} in \mathbb{R}^d is symmetric α -stable if for every $A, B > 0$ there exists a $C > 0$ such that

$$A\mathbf{Z}^{(1)} + B\mathbf{Z}^{(2)} \stackrel{d}{=} C\mathbf{Z}, \tag{16}$$

where $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}$ are independent copies of \mathbf{Z} . Symmetric α -stable random vectors can be represented via their characteristic function, given by [13, Theorem 2.4.3]

$$\mathbb{E}[e^{i\boldsymbol{\theta} \cdot \mathbf{Z}}] = \exp\left(-\int_{\mathbb{S}^{d-1}} \left|\sum_{k=1}^d \theta_k s_k\right|^\alpha \Gamma(ds)\right), \quad \boldsymbol{\theta} \in \mathbb{R}^d, \tag{17}$$

where Γ is the unique symmetric spectral measure on the surface of the d -dimensional unit sphere \mathbb{S}^{d-1} and $0 < \alpha < 2$. A bivariate isotropic sub-Gaussian α -stable random vector \mathbf{Z} corresponds to the case where the spectral measure is uniform and hence admits a characteristic function of the form

$$\mathbb{E}[e^{i\boldsymbol{\theta} \cdot \mathbf{Z}}] = \exp(-\sigma^\alpha |\theta_1 + \theta_2|^\alpha), \quad \theta_1, \theta_2 \in \mathbb{R}. \tag{18}$$

In the special case that $\alpha = 2$, an isotropic sub-Gaussian α -stable random vector is an isotropic Gaussian random vector and, as such, has independent components. For $0 < \alpha < 2$, isotropic sub-Gaussian random vectors have spherical contours; however, the components are dependent. Moreover, for

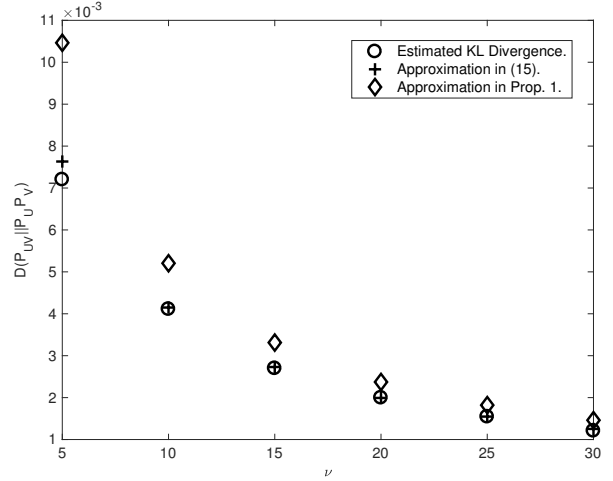


Fig. 1: $D(P_{UV}||P_U P_V)$ for varying values of ν in a bivariate isotropic sub-Gaussian α -stable model with $\alpha = 1.5$ and $\sigma = 1$.

$0 < \alpha < 2$, the marginal distributions of an isotropic sub-Gaussian α -stable random vector are regularly varying with index $-\alpha$ and slowly varying function [13]

$$L(x) = \frac{C_\alpha \sigma^\alpha}{2}, \quad \alpha \neq 1, \quad x \in \mathbb{R}, \tag{19}$$

where

$$C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\pi\alpha/2)}, & \alpha \neq 1 \\ 2/\pi, & \alpha = 1. \end{cases} \tag{20}$$

In Fig. 1, $D(P_{UV}||P_U P_V)$ is estimated for varying choices of $\nu > 0$ when each (X_k, Y_k) are drawn from a bivariate isotropic sub-Gaussian α -stable random distribution compared with $\alpha = 1.5$ and $\sigma = 1$. The estimated value of $D(P_{UV}||P_U P_V)$ is compared with the asymptotic approximations in Prop. 1 and (15). As expected, for sufficiently large ν , both approximations are in good agreement with the estimated value of $D(P_{UV}||P_U P_V)$. For smaller values of ν , the approximation in (15) remains accurate, suggesting that it provides a useful characterization of the error exponent.

As an immediate corollary of Prop. 1, we also have the following characterization of the upper tail dependence.

Corollary 1. *Suppose that P_Y, P_X are regularly varying with index $-a$. Then,*

$$\lambda = \lim_{\nu \rightarrow \infty} \frac{D(P_{UV}||P_U P_V)}{a\bar{F}_X(\nu) \log \nu}. \tag{21}$$

The observation in Corollary 1 provides an information theoretic characterization of the upper tail dependence. Indeed, Corollary 1 suggests that the upper tail dependence can be interpreted as an appropriately scaled distance between the distributions P_{UV} and $P_U P_V$.

IV. TAIL DEPENDENCE IN A SHOT NOISE MODEL

For elliptical distributions with regularly varying marginals, the upper tail dependence admits the characterization in (12). However, unless a copula function is specified, there exists no general characterization for the upper tail dependence of non-elliptical distributions with regularly varying marginals. In this section, we study the upper tail dependence for such a family of distributions arising from a shot noise model.

Consider the random variables,

$$X_i = \sum_{j=0}^{\infty} Z_{ij} r_{ij}^{-\beta}, \quad i = 1, 2, \dots, \quad (22)$$

where $\frac{1}{2} < \beta < \infty$, all weights Z_{ij} are independent with common distribution F_Z satisfying $\mathbb{E}[|Z_{ij}|^{\frac{1}{\beta}}] < \infty$, and $\{r_{ij}\}_{j=1}^{\infty}$, $i = 1, 2$ are dependent homogeneous Poisson point processes, which are constructed as follows. Let Φ be a homogeneous Poisson point process with intensity ξ and let p be the success probabilities. The process $\{r_{ij}\}_{j=1}^{\infty}$ is then obtained by independently thinning Φ with probability p . As such, $\{r_{ij}\}_{j=1}^{\infty}$, $i = 1, 2$ are both homogeneous Poisson point processes with intensity $p\xi$, respectively. It is also clear that the processes are in general dependent as the points $\{r_{ij}\}_{j=1}^{\infty}$, $i = 1, 2$ are drawn from the common process Φ .

The model in (22) is closely related to an interference model arising in wireless communications [10]. More generally, each random variable X_i can be viewed as the long-term behavior of a power-law shot noise process, which has a wide range of applications in physics, biology and finance (see e.g., [14]).

To see that (X_1, X_2) is in general non-elliptical with regularly varying marginals, we have the following theorem, which can be proved using the argument in [10].

Theorem 1. *Suppose that $Z_{ij} \sim \mathcal{N}(0, \sigma^2)$ in (22). Then, the random vector (X_1, X_2) is symmetric $1/\beta$ -stable with characteristic function*

$$\begin{aligned} & \Psi(\theta) \\ &= \exp \left(i(\gamma_1^{1/\beta} |\theta_1^2 + \theta_2^2|^{1/(2\beta)} + \gamma_2^{1/\beta} (|\theta_1|^{1/\beta} + |\theta_2|^{1/\beta})) \right), \end{aligned} \quad (23)$$

where

$$\begin{aligned} \gamma_1 &= \sigma \left(\xi p^2 C_{1/\beta}^{-1} \mathbb{E}[|Z_0|^{1/\beta}] \right)^\beta \\ \gamma_2 &= \sigma \left(\xi p(1-p) C_{1/\beta}^{-1} \mathbb{E}[|Z_0|^{1/\beta}] \right)^\beta \end{aligned} \quad (24)$$

with Z_0 standard normal and $C_{1/\beta}$ as defined in (20).

Indeed, as α -stable distributions are regularly varying and elliptical distributions admit a characteristic function in the form given in (13), it follows immediately from Theorem 1 that the distribution of (X_1, X_2) is non-elliptical in general.

The upper tail dependence of (X_1, X_2) defined in (22) is then given in Theorem 2. Note that the upper tail dependence of the symmetric α -stable random vector considered in Theorem 1 also follows as a special case.

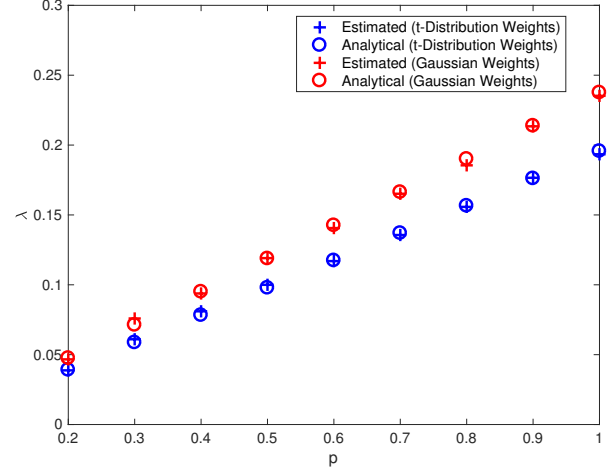


Fig. 2: Upper tail dependence λ for varying values of p with $\beta = 0.7$ and choices of F_Z for the weights $\{Z_{ij}\}$.

Theorem 2. *The upper tail dependence of (X_1, X_2) defined in (22) is given by*

$$\lambda = p \int_0^\infty \bar{F}_Z \left(\left(\frac{2}{\mathbb{E}[|Z_{11}|^{1/\beta}] r} \right)^{-\beta} \right)^2 dr. \quad (25)$$

Proof. See Appendix A. \square

Observe from Theorem 2 that the upper tail dependence of (X_1, X_2) is parameterized by p , which is due to the fact that p determines the probability that X_1, X_2 share a common element of Φ . Moreover, by Theorem 1, (X_1, X_2) is in general non-elliptical, which means that only limited cases (such as when Z_{ij} is Gaussian with $p = 1$) are covered by (12).

To gain insight into the behavior of (25), Fig. 2 plots the upper tail dependence for varying p when each Z_{ij} are t -distributed with a degree of freedom of 4 or standard normal. Observe that the estimated upper tail dependence is in good agreement with (25). Moreover, for non-Gaussian weights, even for $p = 1$, the joint distribution of (X_1, X_2) is not in general elliptical due to the fact that the upper tail dependence differs.

V. CONCLUSIONS

A question arising in dependence testing is the importance of extreme values. For light tailed distributions, such as Gaussian models, extreme values carry little information. On the other hand, for heavy-tailed distributions, including the family of regularly varying distributions, there is significantly more information. In this paper, we studied the error exponent—given by the Kullback-Leibler divergence between the joint and product distributions—for a dependence test based only on data indicating whether or not observations exceed a given value ν . We have established that for large ν the Kullback Leibler divergence is well approximated via an expression

depending on the upper tail dependence. The upper tail dependence has been well characterized for elliptical distributions, but not in general for non-elliptical distributions. To investigate the upper tail dependence in this non-elliptical setting, we derived an expression for a shot noise model, which provided insight into the impact of non-ellipticity on the upper tail dependence.

APPENDIX A
PROOF OF THEOREM 2

We will require the following lemmas.

Lemma 1. Let $\Gamma_1^{-\frac{1}{\alpha}} W_1$ be the first term in the series

$$S = \sum_{i=1}^{\infty} \Gamma_i^{-\frac{1}{\alpha}} W_i, \quad (26)$$

where $0 < \alpha < 2$, $\{\Gamma_i\}_{i=1}^{\infty}$ is a homogeneous Poisson point process with intensity ξ , and $\{W_i\}$ is a sequence of i.i.d. random variables satisfying $\mathbb{E}[|W_i|^\alpha] < \infty$. Suppose that $X \sim S_\alpha(\sigma, 0, 0)$, a symmetric α -stable random variable with scale parameter σ , where

$$\sigma = (\xi C_\alpha^{-1} \mathbb{E}[|W_i|^\alpha])^{1/\alpha}, \quad (27)$$

Then, $X \stackrel{d}{=} S$ and

$$\lim_{x \rightarrow \infty} x^\alpha \Pr(X > x) = \lim_{x \rightarrow \infty} x^\alpha \Pr(\Gamma_1^{-\frac{1}{\alpha}} W_1 > x). \quad (28)$$

Proof. Follows directly from [13, Page 26]. \square

Lemma 2. Let $\{r_{ij}\}$ and $\{Z_{ij}\}$ be defined as in (22). Then,

$$\begin{aligned} & \Pr \left(\sum_{j=1}^{\infty} r_{1j}^{-\beta} Z_{1j} > x, \sum_{j=1}^{\infty} r_{2j}^{-\beta} Z_{2j} > x \right) \\ & \sim \Pr \left(r_{11}^{-\beta} Z_{11} > x, r_{21}^{-\beta} Z_{21} > x \right), \quad x \rightarrow \infty. \end{aligned} \quad (29)$$

Proof. To see that this holds, let $V_i = r_{i1}^{-\beta} W_{i1}$, $i = 1, 2$ and $S_i = \sum_{j=2}^{\infty} r_{ij}^{-\beta} Z_{ij}$, $i = 1, 2$. Observe that for all $s_1, s_2 \in \mathbb{R}$,

$$\Pr(V_1 + s_1 > x, V_2 + s_2 > x) \quad (30)$$

is monotonically decreasing in x . Moreover,

$$\sup_{x \in \mathbb{R}} \left\{ \int_{\mathbb{R}^2} \Pr(V_1 + s_1 > x, V_2 + s_2 > x) f_{S_1 S_2}(s_1, s_2) ds_1 ds_2 \right\} < \infty. \quad (31)$$

As such, Lebesgue's monotone convergence theorem can be applied to swap the limit and integral in

$$\lim_{x \rightarrow \infty} \int_{\mathbb{R}^2} \frac{\Pr(V_1 > x - s_1, V_2 > x - s_2)}{\Pr(X_2 > x)} f_{S_1 S_2}(s_1, s_2) ds_1 ds_2, \quad (32)$$

which, after using Lemma 1 in the denominator, yields the required result. \square

By Lemma 2, to obtain the upper tail dependence it is sufficient to evaluate

$$\lambda = \lim_{u \rightarrow 1} \frac{\Pr \left(r_{11}^{-\beta} Z_{11} > F^{-1}(u), r_{21}^{-\beta} Z_{21} > F^{-1}(u) \right)}{\Pr(X_2 > F^{-1}(u))}, \quad (33)$$

where F is the common distribution function of X_1 and X_2 . To this end, we require the probability that $r_{11} = r_{21}$ conditioned on the event that r_{11} is the k -th point in Φ ; namely,

$$\Pr(r_{11} = r_{21} | k) = (1 - p)^{k-1} p. \quad (34)$$

Similarly, the probability that r_{11} is the k -th point in Φ is given by

$$\Pr(k) = (1 - p)^{k-1} p. \quad (35)$$

Let $f_{r_{11}}(r_{11} | k)$ denote the probability density of r_{11} conditioned on the event it is the k -th point in Φ . Under this conditioning, r_{11} is Gamma distributed and

$$f_{r_{11}}(r_{11} | k) = \frac{1}{(k-1)!} \xi^k r_{11}^{k-1} e^{-\xi r_{11}}. \quad (36)$$

We then have

$$\begin{aligned} & \frac{\Pr \left(r_{11}^{-\beta} Z_{11} > F^{-1}(u), r_{21}^{-\beta} Z_{21} > F^{-1}(u) \right)}{\Pr(X_2 > F^{-1}(u))} \\ & \sim \frac{1}{\Pr(X_2 > F^{-1}(u))} \sum_{k=1}^{\infty} ((1-p)^{k-1} p)^2 \\ & \cdot \int_0^{\infty} dr_{11} \frac{1}{(k-1)!} \xi^k r_{11}^{k-1} e^{-\xi r_{11}} \\ & \cdot \Pr(r_{11}^{-\beta} Z_{11} > F^{-1}(u), r_{11}^{-\beta} Z_{21} > F^{-1}(u) | r_{11} = r_{21}, r_{11}, k), \end{aligned} \quad (37)$$

as $u \rightarrow 1$. This follows from the fact that the term arising from conditioning on the event $\{r_{11} \neq r_{21}\}$ has a negligible contribution as $u \rightarrow 1$ since $r_{11}^{-\beta} Z_{11}$ and $r_{21}^{-\beta} Z_{21}$ are independent in this case.

As Z_{11}, Z_{21} are independent, we have

$$\begin{aligned} & \Pr(r_{11}^{-\beta} Z_{11} > F^{-1}(u), r_{11}^{-\beta} Z_{21} > F^{-1}(u) | r_{11} = r_{21}, r_{11}, k) \\ & = \bar{F}_Z \left(\frac{F^{-1}(u)}{r_{11}^{-\beta}} \right)^2, \end{aligned} \quad (38)$$

since Z_{11} and Z_{21} have a common distribution F_Z .

From Lemma 1 and (20), we have

$$F_i^{-1}(u) \sim \left(\frac{2(1-u)}{\xi p \mathbb{E}[|Z_{i1}|^{\frac{1}{\beta}}]} \right)^{-\beta}. \quad (39)$$

After applying the change of variables $r = r_{11}(1-u)^{-1}$ in (37), we then have,

$$\begin{aligned} & \lim_{u \rightarrow 1} \frac{\Pr \left(r_{11}^{-\beta} Z_{11} > F^{-1}(u), r_{21}^{-\beta} Z_{21} > F^{-1}(u) \right)}{\Pr(X_2 > F^{-1}(u))} \\ & = p^2 \xi \int_0^{\infty} \bar{F}_Z \left(\left(\frac{2}{\xi p \mathbb{E}[|Z_{11}|^{\frac{1}{\beta}}]} r \right)^{-\beta} \right)^2 dr, \end{aligned} \quad (40)$$

from which the result readily follows.

REFERENCES

- [1] T. Berrett, I. Kontoyiannis, and R. Samworth, “Optimal rates for independence testing via u -statistic permutation tests,” *arXiv:2001.05513*, 2020.
- [2] I. Kojadinovic and M. Holmes, “Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process,” *J. Mult. Anal.*, vol. 100, pp. 1137–1154, 2009.
- [3] S. Filippi and C. Holmes, “A Bayesian nonparametric approach to testing for dependence between random variables,” *Bayesian Analysis*, vol. 12, no. 4, pp. 919–938, 2017.
- [4] Y. Polyanskiy, H. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [5] H. Poor, *An Introduction to Signal Detection and Estimation*. Springer Science & Business Media, 2013.
- [6] T. Han and S. Amari, “Statistical inference under multiterminal data compression,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, Oct.
- [7] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.
- [8] S. Resnick, *Extreme values, regular variation and point processes*. Springer, 2013.
- [9] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modeling Extremal Events: For Insurance and Finance*. Springer Science & Business Media, 2013.
- [10] C. Zheng, M. Egan, L. Clavier, T. Pedersen, and J.-M. Gorce, “Linear combining in dependent α -stable interference,” in *IEEE International Conference on Communications (ICC)*, 2020.
- [11] R. Nelson, *An Introduction to Copulas*. New York, NY: Springer-Verlag, 1999.
- [12] S. Demarta and A. McNeil, “The t -copula and related copulas,” *International Statistical Review*, vol. 73, no. 1, pp. 111–129, 2005.
- [13] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*. New York, NY: CRC Press, 1994.
- [14] S. Lowen and M. Teich, “Power-law shot noise,” *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1302–1318, Nov. 1990.