

Copulas and Dependency Measures for Multivariate Linnik Distributions

Alexandru Belu¹, Wesley Maddox² & Wojbor A. Woyczynski³

¹ Bloomberg, LP., Financial Software Engineering, New York, NY, USA

² Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA

³ Department of Mathematics, Applied Mathematics and Statistics, and Center for Stochastic and Chaotic Processes in Science and Technology, Case Western Reserve University, Cleveland, OH 44122, USA

Correspondence: Wojbor A. Woyczynski, Department of Mathematics, Applied Mathematics and Statistics, and Center for Stochastic and Chaotic Processes in Science and Technology, Case Western Reserve University, Cleveland, OH 44122, USA

Received: September 24, 2018 Accepted: October 22, 2018 Online Published: October 31, 2018

doi:10.5539/ijsp.v7n6p154 URL: https://doi.org/10.5539/ijsp.v7n6p154

Abstract

We begin this paper by introducing the Linnik distributions in both the univariate and multivariate case. An overview of simulation methods and two estimation procedures for the multivariate Linnik distribution are presented. Experiments demonstrating the accuracy of the procedures are also included. Then a novel multivariate Linnik copula is derived. The primary focus of this part of work is on simulation and estimation procedures for this copula, applying existing algorithms for simulation and estimation procedures for the multivariate Linnik distribution derived in the prior section. Several theoretical properties of the copula in relation to different dependence metrics are derived.

Keywords: Copulas, α -stable processes, α -Linnik processes, estimation of copulas, dependence metrics

1. Introduction: Linnik Distributions

In this section we recall the basic classical concepts related to Lévy processes, and infinitely divisible distributions. Two special cases, one-dimensional and multivariate, are considered: the α -stable processes, and α -Linnik processes. Then the fundamental properties of the two classes are provided and compared with each other. In particular, the differences in the tail behavior of their distributions, and the fractional moment properties are explained. A study of the nonlinear evolution equations driven by the infinitesimal generators of Linnik processes can be found in Gunaratnam and Woyczynski(2015).

1.1 Preliminaries on General One-Dimensional Lévy Processes and Infinitely Divisible Distributions

A stochastic process X_t , $t \geq 0$, is called a *Lévy process* (see, e.g, Bertoin (1996), Samorodnitsky and Taqqu (1994), and Sato (1999)) if :

- (i) $X_0 = 0$, with probability 1,
- (ii) X_t has independent increments, i.e., for any $n \in \mathbb{N}$, and any $0 < t_1 < t_2 < \dots < t_n$, the random variables $X_{t_2} - X_{t_1}$, $X_{t_3} - X_{t_2}$, \dots , $X_{t_n} - X_{t_{n-1}}$ are independent,
- (iii) X_t has stationary increments, i.e., for any $t > s \geq 0$, $X_t - X_s$ is equal in distribution to X_{t-s}
- (iv) The trajectories of X_t are right continuous with left limits, with probability 1.

The 1-D distributions of the Lévy processes are *infinitely divisible* (ID) which, in terms of the characteristics functions, can be expressed as the obvious equality:

$$\phi_X(\xi, t) = \mathbb{E} \exp(i\xi X(t)) = [\phi_X(\xi, t/n)]^n, \quad \xi \in \mathbb{R}. \quad (1.1)$$

for every $n \geq 1, t \geq 0$. A more detailed description of the structure of characteristic functions of infinitely divisible distributions is given by the following classical *Lévy-Khinchine Representation Theorem*:

THEOREM 2.1. *A random variable X has an infinitely divisible distribution if, and only if, there exist $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$, and a nonnegative measure Λ on $\mathbb{R} \setminus \{0\}$ satisfying the condition $\int_{\mathbb{R}} (1 \wedge |x|^2) \Lambda(dx) < \infty$, such that its characteristic function is of the form,*

$$\phi_X(\xi) := \mathbb{E}[e^{i\xi X}] = e^{\psi(\xi)}, \quad (1.2)$$

where the characteristic exponent

$$\psi(\xi) = i\mu\xi - \frac{(\sigma\xi)^2}{2} + \int_{\mathbb{R} \setminus \{0\}} (e^{i\xi x} - 1 - i\xi x \mathbf{1}_{|x| < 1}) \Lambda(dx). \quad (1.3)$$

The triplet (μ, σ, Λ) is called the *characteristic triplet* of X , $\psi(u)$ – the characteristic exponent of X , $\mu \in \mathbb{R}$ – the drift coefficient, $\sigma > 0$ – the Gaussian, or diffusion coefficient, and Λ – the Lévy measure of X . The Lévy measure describes the “intensity” of jumps of a certain size of a Lévy process.

1.2 One-Dimensional α -Stable Distributions

The self-similar (symmetric) Lévy probability distributions form the family of the classical α -stable distributions (denoted $S(\alpha, c)$ in this paper) with the characteristic functions of the form

$$\phi_S(\xi; \alpha, c) = \exp(-|c\xi|^\alpha), \quad (1.4)$$

where the constant $0 < \alpha \leq 2$, is called the index of stability, and c is the scale parameter. In general, probability density functions (PDFs), $f_S(x; \alpha, c)$, of the α -stable random variables do not have a clean closed-form representations in terms of elementary functions. Apart from the obvious Gaussian case of $\alpha = 2$, we would like to mention the Cauchy case, $\alpha = 1$, in which case the PDF is of the form,

$$f_S(x; 1, c) = \frac{c}{\pi(c^2 + x^2)}. \quad (1.5)$$

The Lévy measure Λ_S of the distribution of $S(\alpha, c)$ is absolutely continuous with respect to the Lebesgue measure, and its density is given by the expression,

$$\lambda_S(x) = \frac{\Lambda_S(dx)}{dx} = \frac{C}{|x|^{\alpha+1}}, \quad (1.6)$$

where the easily determined constant $C = c^\alpha (2 \int_0^\infty (\cos v - 1) v^{-(1+\alpha)} dv)^{-1}$.

1.3 One-Dimensional α -Linnik Distributions and Their Basic Properties

The univariate α -Linnik distribution (denoted $L(\alpha, \gamma)$ in this paper) with index $\alpha \in (0, 2]$, and the scale parameter $\gamma > 0$, is defined by its characteristic function (see Linnik (1953))

$$\phi_L(\xi; \alpha, \gamma) = \frac{1}{1 + |\gamma\xi|^\alpha}, \quad t \in \mathbb{R}, \quad \gamma > 0. \quad (1.7)$$

Let us denote its density function by $f_L(x; \alpha, \gamma)$. In the extreme case $\alpha = 2$, and $\gamma = 1$, the Linnik distribution corresponds to the standard Laplace (two-sided exponential) distribution with the density function,

$$f(x; 2, 1) = \frac{1}{2} e^{-|x|/2}; \quad x \in \mathbb{R}. \quad (1.8)$$

By inverting the characteristic function $\phi_L(x; \alpha, 1)$ one can immediately see (1995, 2001) that $f_L(x; \alpha, 1) = f_L(-x; \alpha, 1)$, and

$$\begin{aligned} f_L(x; \alpha, 1) &= \frac{1}{\pi} \int_0^\infty \frac{1}{1+t^\alpha} \cos(xu) du \\ &= \frac{\sin(\pi\alpha/2)}{\pi} \int_0^\infty \frac{t^\alpha e^{-t|x|} du}{1+t^{2\alpha} + 2t^\alpha \cos(\pi\alpha/2)}. \end{aligned}$$

Note that, for $1 < \alpha \leq 2$, $f_L(0; \alpha, 1) < \infty$, but, for $0 < \alpha \leq 1$, the density has a singularity at the origin, $f_L(0; \alpha, 1) = \infty$. Also, the absolute first moment is finite only for $1 < \alpha \leq 2$, and the variance is well defined only in the case $\alpha = 2$. The distribution is unimodal, and Devroye (1990) provides an elegant method of generating Linnik random variables. A few sample PDFs are shown in Fig. 2.1, below.

For $\alpha \in (0, 2)$, the α -Linnik distributions, like the corresponding α -stable distributions, have ‘fat tails’ (no longer exponential). More precisely, they have the following asymptotic behavior at infinity:

THEOREM 2.2. (see Kotz, Kozubowski and Hayfavi (1995)) *For any $\alpha \in (0, 2)$,*

$$f_L(x; \alpha, 1) \sim \frac{1}{\pi} \sum_{k=1}^{\infty} \left\{ (-1)^{k-1} \Gamma(1 + \alpha k) \sin(\pi\alpha k/2) \right\} |x|^{-1-\alpha k}, \quad \text{as } x \rightarrow \infty, \quad (1.9)$$

or, equivalently,

$$f_L(x; \alpha, 1) = \frac{1}{\pi} \sum_{k=1}^n \left\{ (-1)^{k-1} \Gamma(1 + \alpha k) \sin(\pi \alpha k / 2) \right\} |x|^{-1-\alpha k} + R_{n,\alpha}(x), \quad (1.10)$$

where

$$|R_{n,\alpha}(x)| \leq \frac{\Gamma(1 + \alpha(n+1))}{\pi \sin(\pi \alpha / 2)} |x|^{-1-\alpha(n+1)}. \quad (1.11)$$

Consequently, for any $\alpha \in (0, 2)$, the density $f_L(x; \alpha, 1)$ decreases at ∞ at the rate of the power function $x^{-(1+\alpha)}$. More precisely,

$$f_L(x; \alpha, 1) \sim \frac{1}{\pi} \{ \Gamma(1 + \alpha) \sin(\pi \alpha / 2) \} |x|^{-(1+\alpha)}, \quad \text{as } x \rightarrow \infty. \quad (1.12)$$

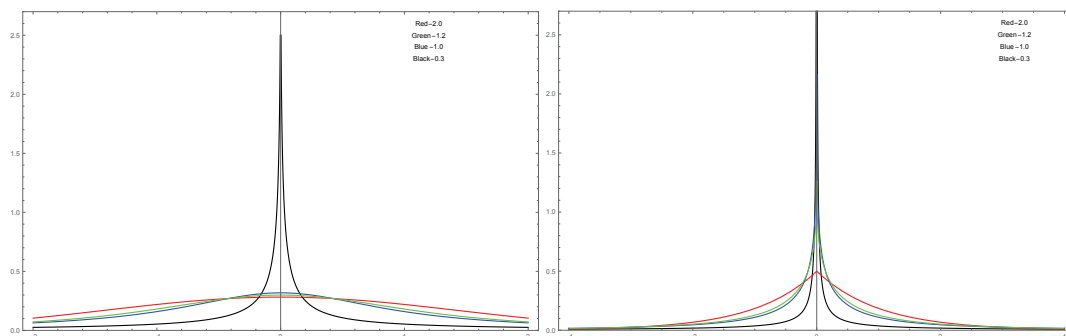


Figure 1.1. PDFs of α -stable (left), and α -Linnik distributions (right) for $\alpha = 2$ (red), 1.25 (green), 1.0 (blue), and 0.3 (grey). Note the singularity at zero for $\alpha = 1$, and $\alpha = 0.3$, for the α -Linnik distribution.

The importance of α -stable distributions comes from their universality as approximants of the distributions of sums of a fixed but large numbers of independent, identically distributed random variables. Indeed, the fundamental central limit theorem states that if $X_1, X_2, \dots, X_n, \dots$ are independent, identically distributed (i.i.d.) random variables then their sums $S_n = X_1 + X_2 + \dots + X_n$ converge, as $n \rightarrow \infty$, in distribution (after some rescaling) to an α -stable distribution for some $\alpha \in (0, 2]$. This fact is directly related to the “stability” property: if X_1, X_2, \dots, X_n are symmetric α -stable themselves then $Y = n^{-1/\alpha}(X_1 + X_2 + \dots + X_n)$ has the same distribution as each of the X_i ’s. A method of simulating stable random variables can be found in Chambers, Mallows and Stuck (1976).

The α -Linnik distributions have a similar stability property, but under random geometric summation. Indeed (see also Kozubowski (1994)), we have the following simple result:

PROPOSITION 2.1. *If X, X_1, X_2, \dots are i.i.d. random variables and N is an independent of X_1, X_2, \dots random variable with the geometric distribution with mean $1/p$, $0 < p < 1$, that is, $\mathbf{P}(N = n) = p(1-p)^{n-1}$, $n = 1, 2, \dots$, then the following two statements are equivalent:*

(i) X is α -stable with respect to geometric summation, i.e.,

$$p^{1/\alpha} \sum_{i=1}^N X_i \stackrel{d}{=} X,$$

(ii) X has the α -Linnik distribution.

Proof. The verification is straightforward via the characteristic function:

$$\begin{aligned} & \mathbf{E} \exp[i\xi p^{1/\alpha}(X_1 + X_2 + \dots + X_N)] \\ &= \sum_{n=1}^{\infty} \mathbf{E}(\exp[i\xi p^{1/\alpha}(X_1 + X_2 + \dots + X_N)] | N = n) p(1-p)^{n-1} \\ &= \sum_{n=1}^{\infty} \left(\frac{1}{1 + |\gamma \xi p^{1/\alpha}|^\alpha} \right)^n p(1-p)^{n-1} = \frac{1}{1 + |\gamma \xi|^\alpha}. \quad \text{QED} \end{aligned}$$

Two elementary, but crucial observations are in order here:

Remark 2.1. The α -Linnik distributions are in the domain of attraction of the α -stable distribution, that is, if $Z_n = n^{-1/\alpha}(X_1 + X_2 + \cdots + X_n)$ are i.i.d. and $X_i \sim L(\alpha, \gamma)$ for all i , then the distribution of Z_n converges to an α -stable distribution as n tends to infinity. Indeed, working again with characteristic functions, we get

$$\lim_{n \rightarrow \infty} \mathbf{E}(\exp(i\xi Z_n)) = \lim_{n \rightarrow \infty} (1 + |\xi\gamma|^\alpha/n)^{-n} = \exp(-|\gamma\xi|^\alpha).$$

Remark 2.2. If Z is a standard exponential random variable, and $Y \sim S(\alpha, c)$ is an independent α -stable random variable then

$$X = Z^{1/\alpha} Y \quad (1.13)$$

is a Linnik random variable with the distribution $L(\alpha, \gamma)$, and $\gamma = c$, see Devroye (1990). Indeed,

$$\mathbf{E} e^{i\xi Z^{1/\alpha} Y} = \int \int e^{i\xi z^{1/\alpha} y} f_Y(y) e^{-z} dy dz = \int_0^\infty e^{-z(1+|c\xi|^\alpha)} dz = \frac{1}{1+|c\xi|^\alpha}.$$

Remark 2.3. If N is a normal $N(0, \gamma)$ random variable, and $M_{\alpha/2}$ is an independent Mittag-Leffler random variable, then

$$X = M_{\alpha/2}^{1/2} N \quad (1.14)$$

is a Linnik random variable with the distribution $L(\alpha, \gamma)$. Recall, that the Mittag-Leffler distribution with parameter α is typically defined in terms of its distribution function

$$F(x; \alpha) = 1 - E(-x^\alpha; \alpha),$$

where

$$E(z; \alpha) := \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + 1)}$$

is the standard Mittag-Leffler special function.

1.4 Comparison of Tails and Lower-order Fractional Moments of α -Stable and α -Linnik Distributions

For $0 < \alpha < 2$, the tail behavior of the distribution of $Y \sim S(\alpha, c)$ is as follows (see Samorodnitsky and Taqqu (1994)):

$$\lim_{y \rightarrow \infty} y^\alpha \mathbf{P}(|Y| > y) = C_\alpha c^\alpha, \quad (1.15)$$

where

$$C_\alpha = \frac{\sin(\pi\alpha/2)\Gamma(\alpha)}{\pi}, \quad (1.16)$$

In the α -Linnik case we have a similar result.

PROPOSITION 2.2. If $X \sim L(\alpha, \gamma)$, $0 < \alpha < 2$, then

$$\lim_{y \rightarrow \infty} y^\alpha \mathbf{P}(|X| > y) = C_\alpha \gamma^\alpha. \quad (1.17)$$

Proof: Indeed, in view of Remark 2.2,

$$\begin{aligned} \lim_{x \rightarrow \infty} x^\alpha \mathbf{P}(|X| > x) &= \lim_{x \rightarrow \infty} x^\alpha \mathbf{P}(|Y| > xZ^{-1/\alpha}) \\ &= \lim_{x \rightarrow \infty} x^\alpha \int \int_{\{|y| > xz^{-1/\alpha}\}} f_Y(y) e^{-z} dy dz \\ &= \int_0^\infty \left(\lim_{x \rightarrow \infty} (xz^{-1/\alpha})^\alpha \int_{\{|y| > xz^{-1/\alpha}\}} f_Y(y) dy \right) ze^{-z} dz \\ &= C_\alpha \gamma^\alpha \int_0^\infty ze^{-z} dz = C_\alpha \gamma^\alpha. \quad \text{QED} \end{aligned}$$

Given the above asymptotic results (see also Theorem 2.2), both, the α -Linnik and α -stable random variables with $0 < \alpha < 2$ have finite moments of fractional order less than α ; the higher-order moments are infinite. More precisely, if $Y \sim S(\alpha, c)$, and $0 < p < \alpha$, then (see Samorodnitsky and Taqqu (1994))

$$\mathbf{E}|Y|^p = \frac{2^{p+1} c^p \Gamma\left(\frac{p+1}{2}\right) \Gamma\left(-\frac{p}{\alpha}\right)}{\alpha \sqrt{\pi} \Gamma\left(-\frac{p}{2}\right)}. \quad (1.18)$$

On the other hand, for the α -Linnik distributions we have the following result:

PROPOSITION 2.3. Let $X \sim L(\alpha, \gamma)$, with $0 < \alpha < 2$. Then, for every p , $0 < p < \alpha$,

$$\mathbf{E}|X|^p = \frac{2^{p+1} p \gamma^p \Gamma(\frac{p+1}{2}) \Gamma(-\frac{p}{\alpha}) \Gamma(\frac{p}{\alpha})}{\alpha^2 \sqrt{\pi} \Gamma(-\frac{p}{2})}. \quad (1.19)$$

Proof: In view of Remark 2.2,

$$\mathbf{E}|X|^p = \mathbf{E}Z^{p/\alpha} \cdot \mathbf{E}|Y|^p = \int_0^\infty z^{p/\alpha} e^{-z} dz \cdot \mathbf{E}|Y|^p = \Gamma\left(\frac{p}{\alpha} + 1\right) \mathbf{E}|Y|^p,$$

which gives (2.19) as a result of (2.18), and the identity $\Gamma(x+1) = x\Gamma(x)$. QED

Remark 2.4. Thus the α -stable and α -Linnik distributions have exactly matching probability distribution tails if $c = \gamma$. However, the exactly matching tails of $S(\alpha, c)$ and $L(\alpha, \gamma)$ in the case $c = \gamma$ do not imply that their fractional moments match. To assure the equality of the p -th order fractional lower order moment of the α -Linnik distribution $L(\alpha, \gamma)$, and the α -stable distribution $S(\alpha, c, \beta, \mu)$, which may be also useful in simulations, their parameters have to be related via the following equality

$$\frac{p \gamma^p \Gamma(p/\alpha)}{\alpha} = c^p, \quad 0 < p < \alpha. \quad (1.20)$$

For example, if $\alpha = 3/2$, then to assure the equality of the first-order moments we must have $c = 0.902745\gamma$. For more information on analytic and asymptotic properties of Linnik probability densities see Kotz, Kozubowski and Hayfavi (1995), Kotz, Kozubowski and Podgorski (2001), and Kozubowski and Panorska (1999).

1.5 Lévy Measure of the Linnik Distribution

There is an obvious relationship between the characteristic function $\phi_S(\xi; \alpha, 1)$ of the symmetric stable distribution, and the characteristic function $\phi_L(\xi; \alpha, 1)$ of the Linnik distribution:

$$\phi_L(\xi) = \frac{1}{1 - \log \phi_S(\xi)}. \quad (1.21)$$

This fact permits expressing the density of the Lévy measure of Λ_L of the α -Linnik distribution in terms of the α -stable distribution (see, also Kozubowski and Rachev (1999)):

$$\frac{\Lambda_L(dx)}{dx} = \frac{\alpha}{2|x|} \mathbf{E} \exp\left(-\left|\frac{x}{X}\right|^\alpha\right), \quad (1.22)$$

where $X \sim S(\alpha, 1)$.

Finally, using the tail asymptotics (2.14) of the $S(\alpha, c)$, one obtains the following asymptotics for the Lévy measure of the α -Linnik distribution:

$$\lim_{x \rightarrow \infty} x^\alpha \Lambda_L([x, \infty)) = \frac{1}{2\Gamma(1-\alpha) \cos(\pi\alpha/2)}; \quad 0 < \alpha < 2. \quad (1.23)$$

Remark 2.5. Generalized Linnik Distributions. Since Linnik distributions belong to the much larger class of geometric stable distributions (which also include Mittag-Leffler distributions), various different generalizations of the Linnik distribution exist. A commonly used generalization, called the *asymmetric Linnik distribution*, is defined by the characteristic function,

$$\phi_{GL}(\xi; \alpha, \theta) = \frac{1}{1 + |\xi|^\alpha \exp(i\theta \operatorname{sgn}(\xi))} \quad (1.24)$$

1.6 Multivariate Linnik Distributions

Finally, we turn to the description of the multivariate d -dimensional Linnik distributions which will be the main object of investigation in the remainder of this paper. Despite the potential usefulness of these models in active applied research, not much effort has been made so far to develop methods for simulation and parameter estimation of these distributions.

Following Anderson (1992) we define a multivariate Linnik distribution $L(\alpha, \Omega)$ via its characteristic function,

$$\phi_L(\xi) = \frac{1}{1 + (\xi^T \Omega \xi)^{\alpha/2}}, \quad (1.25)$$

where the transpose $\boldsymbol{\xi}^T = (\xi_1, \dots, \xi_d)$. The two parameters here are: α , which represents the thickness of the tail of the distribution, serving the same purpose as in univariate α -stable, and α -Linnik distributions, and the matrix Ω . Here, Ω is assumed to be symmetric and positive definite, much like a covariance matrix that is used to characterize a multivariate normal distribution. In the one dimensional case, this characterization reduces to Eq. (2.7), where $\gamma = \Omega^2$.

An integral expression for the PDF corresponding to this characteristic function is of the form (see Ostrovskii (1995), and Kozubowski (2000)),

$$g_L(\mathbf{y}) = \frac{2 \sin(\pi\alpha/2) (\sqrt{2\mathbf{y}^T \Omega^{-1} \mathbf{y}})^{1-d/2}}{\pi^{1+d/2} |\Omega|^{1/2}} \int_0^\infty \frac{K_{1-d/2}(u \sqrt{\mathbf{y}^T \Omega^{-1} \mathbf{y}}) u^{\alpha+d/2}}{1 + u^{2\alpha} + 2u^\alpha \cos(\pi\alpha/2)} du, \quad (1.26)$$

where K_ν is the Bessel function of the third kind. When $\alpha = 2$, and $d = 2$, this aligns with the standard bivariate distribution function for the Laplace distribution. However, due to its complexity, this expression is typically of limited usefulness.

Remark 2.5. In analogy to the one-dimensional representation of Linnik distribution described in Remark 2.3 we also have the following multivariate representation: If \mathbf{N} is a normal $N(0, \Omega)$ random vector, and $M_{\alpha/2}$ is an independent Mittag-Leffler random variable, then

$$\mathbf{X} = M_{\alpha/2}^{1/2} \mathbf{N} \quad (1.27)$$

is a Linnik random vector with the distribution $L(\alpha, \Omega)$.

2. Copulas and Simulation of Linnik Distributions

2.1 Basic Facts About Copulas

For a random vector (X_1, X_2, \dots, X_d) with continuous marginal cumulative distribution functions (CDFs), $F_i(x) = P[X_i \leq x]$ the copula function is defined as the joint CDF of the random vector

$$(U_1, U_2, \dots, U_d) := (F_1(X_1), F_2(X_2), \dots, F_d(X_d)),$$

that is

$$C(u_1, u_2, \dots, u_d) = P[U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d].$$

So, the copula function contains complete information about the dependence structure between the components of the random vector (X_1, X_2, \dots, X_d) . The classical 1959 Sklar's Theorem (see, e.g., Nelsen (2006)) states that every multivariate cumulative distribution function,

$$F(x_1, \dots, x_d) = P[x_1 \leq x_1, \dots, X_d \leq x_d],$$

can be expressed in terms of its marginal CDFs, $F_i(x_i)$, $i = 1, \dots, d$, by the formula

$$F(x_1, \dots, x_d) = C(F_1(X_1), F_2(X_2), \dots, F_d(X_d)). \quad (2.1)$$

Nice and concise proofs of this result, in both the bivariate and multidimensional cases, are given in Nelsen (2006). Since (U_1, U_2, \dots, U_d) have uniform marginal distributions on the unit interval $I = [0, 1]$, a copula function in d -dimensions is a function mapping $[0, 1]^d$ into the unit interval $[0, 1]$ with the following basic properties:

- (i) For every vector-valued $\mathbf{u} \in I^d$, if at least one $u_i = 0$, then $C(\mathbf{u}) = 0$. This is the so-called *groundedness property* of copulas.
- (ii) Similarly, if all coordinates $u_i = 1$ except for possibly u_k , then $C(\mathbf{u}) = u_k$.
- (iii) The copula function is uniformly continuous on its domain, that is for every \mathbf{u} and $\mathbf{v} \in I^n$

$$|C(\mathbf{v}) - C(\mathbf{u})| \leq \sum_{k=1}^n |v_k - u_k|.$$

- (iv) Additionally, the marginal derivatives, $C'_k(u)$, all exist and are equal to u , ie

$$C'_k(u) = u \quad \forall u \in [0, 1].$$

Taken together, these properties show that the copula function, $C(\mathbf{u})$ is grounded and increasing in every dimension (termed n -increasing) (Nelsen (2006)).

While (3.1) defines the joint distribution function in terms of the marginals and the copula function, the inverse also holds. Taking the generalized inverse of the marginal distributions gives the copula function in terms of the marginals and the joint distribution,

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)). \quad (2.2)$$

It is worthwhile to remember this relationship, as it is important in both sampling and estimation procedures for copulas (Nelsen (2006)).

Similarly, the joint probability density function (PGF), $f(x_1, \dots, x_d)$, is also well-defined in terms of copulas. By taking successive partial derivatives $\partial/\partial x_i$ of the copula function, we can define the copula density function as

$$c(x_1, \dots, x_d) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} C(x_1, \dots, x_d).$$

Thus, a simple application of the chain rule of differentiation gives the formula for the joint PDF:

$$f(x_1, \dots, x_d) = c(x_1, \dots, x_d) \prod_{i=1}^d f(x_i). \quad (2.3)$$

If the copula density is well-defined and easy to use, this gives a powerful tool for maximum likelihood estimation and sampling. However, in many cases, despite its existence finding it analytically may be intractable.

2.2 Examples

In this subsection, several examples of copula functions are given to make the preceding discussion on copulas less abstract.

Example 3.2.1. Independence Copula. The most basic copula is the independence copula, which codifies the relationship between independent random variables. Its obvious copula function is

$$C(u_1, \dots, u_d) = \prod_{i=1}^d u_i = u_1 \dots u_d. \quad (2.4)$$

Example 3.2.2. Gaussian Copulas. A more widely used and useful copula function is the Gaussian copula, which encodes a Gaussian relationship between distributions of random variables. Its copula function is given by

$$C(u_1, \dots, u_d; \Sigma) = \phi_{0, \Sigma}(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)), \quad (2.5)$$

where $\phi_{0, \Sigma}$ is the distribution function of the multivariate zero-mean Gaussian vector with covariance matrix Σ , and ϕ^{-1} is the quantile function of the standard normal 1-D distribution. This construction is efficient for both simulation and estimation, as the only parameters which must be estimated for the copula are just the rank-based correlations between the marginal distributions (easily estimated).

Through manipulation of the Gaussian density function, it is possible to get the copula density function (see Zezula (2009)). Since the multivariate density function of the Gaussian distribution with zero mean and covariance matrix Σ is of the form,

$$f(x_1, \dots, x_d; \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2} \prod_{i=1}^d \sigma_{ii}} \exp\left(-\frac{1}{2} \mathbf{u}^T \Sigma^{-1} \mathbf{u}\right),$$

where $u_i = x_i/\sigma_{ii}$, it can be rewritten in the form,

$$f(x_1, \dots, x_d; \Sigma) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{u}^T (R^{-1} - I_d) \mathbf{u}\right) \prod_{i=1}^d \frac{1}{\sigma_i} \phi(u_i),$$

where $R_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ (i.e., R is the correlation matrix corresponding to Σ). In view of Eq 2.3 the copula density function in this case is simply of the form,

$$c(x_1, \dots, x_d; \Sigma) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{u}^T (R^{-1} - I_d) \mathbf{u}\right). \quad (2.6)$$

This density function is highly useful as it can be used for either simulation or analytic purposes.

The Gaussian copula is part of a larger class of elliptical copula functions, which follow the same type of structure (and thus simulation and estimation algorithms). Additionally, the ease of simulation and estimation for these copulas have led them to become widely used across mathematical finance (where they were the core of the original CreditMetrics model (see Li (2000)) and are increasingly used in the machine learning literature (see Rey (2015), Wilson and Ghahramani (2010)).

Example 3.2.3. Elliptical Copulas. A random vector \mathbf{X} is said to be elliptically distributed if its characteristic function $\phi(\xi)$ satisfies the following functional equation

$$\phi_{\mathbf{X}-\mu}(\xi) = \psi(\xi^T \Sigma \xi),$$

for some location parameter μ , some nonnegative definite matrix Σ , and some scalar function ψ . In terms of the density functions an elliptical distribution can be written in the form,

$$f_{\mathbf{X}}(\mathbf{x}) = ag((\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)),$$

where a represents a scale parameter, g is a scalar function, μ is the median vector, and Σ is a positive definite matrix.

Alternatively, a random d -dimensional vector \mathbf{X} has an elliptical distribution if it admits a representation

$$\mathbf{X} = \mu + A^T R \mathbf{S}$$

where $\mu \in \mathbf{R}^d$, A is a $k \times d$ -matrix, and R is a positive random variable (random radius) transforming the k -dimensional random vector \mathbf{S} which is uniformly distributed on the unit sphere.

An elliptical copula is thus defined as

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)),$$

where F is the distribution function of the related elliptical distribution and

$$F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)$$

are the corresponding univariate quantile functions (see Mai and Scherer (2012) for more details).

Example 3.2.4. Empirical Copulas. The empirical copula is a parameter-less copula function that can be defined directly from a sample of random variables (see Nelsen (2006)). The following definition formulated just in the two-dimensional case can easily be extended to d -dimensions.

The empirical copula of a sample of size n of a continuous bivariate distribution is defined as

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\#\{(x, y) : x \leq x_{(i)}, y \leq y_{(j)}\}}{n},$$

where $x_{(i)}$ and $y_{(j)}$ represent the i -th and j -th order statistics from the bivariate sample.

Alternatively and equivalently (see Ruschendorf (2009)), the empirical copula can be described in terms of the univariate empirical distribution function, which is simply

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x).$$

If we let $F_n(x_i) = U_i$, we clearly know that $U_i \in [0, 1]$, then the empirical copula function is simply

$$C_n(u_i, \dots, u_d) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(U_j \leq u_j).$$

It is worth noting that the empirical copula function has been shown to be a consistent estimator of the true copula function, and that it converges uniformly to the true distribution function. Additionally, the error of replacing the true copula by the empirical copula decreases exponentially to zero (see Poczos, Ghahramani and Schneider (2012)), so that, for any $\epsilon > 0$,

$$P[\sup_{\mathbf{x} \in \mathbb{R}^d} \|C(\mathbf{u}) - C_n(\mathbf{u})\|_2 > \epsilon] \leq 2d \exp\left\{-\frac{n\epsilon^2}{d}\right\}, \quad (2.7)$$

where \mathbf{u} is just (u_1, \dots, u_d) written in vector notation. As several standard measures of dependence can be written in terms of the related copula function, the above result shows that numerical calculations using the empirical copula will provide strong estimates of these measures of dependence.

2.3 Multivariate Simulation

Several different methods for sampling from copula distributions exist. Drawing samples in the bivariate case is often accomplished conditionally (see Chapter 1 of Mai and Scherer (2012)), while in higher dimensions, algorithms typically are dependent on the class that the copula belongs to (see Chapters 2-5 of Mai and Scherer (2012)). In this subsection, only one simulation algorithm is provided due to its usefulness in the rest of the paper.

Happily, work from the theory of one-dimensional geometrically stable distributions can be easily extended, so that it is possible to simulate multivariate Linnik distributions (see Kozubowski (2000)). In one dimension, another equivalence holds for the Linnik distribution,

$$Y_{\alpha,\gamma} \stackrel{d}{=} \sqrt{M_{\alpha/2}} X$$

where M_p is a Mittag-Leffler random variable and $X \sim N(0, \gamma)$. This relation is due to the geometric stability of Linnik random variables and through manipulations of the mixture representation of Linnik random variables. Since this holds, it easily gives rise to a generator for the multivariate Linnik distribution, $Y \sim L(\alpha, \Sigma)$,

$$Y_{\alpha,\Sigma} \stackrel{d}{=} \sqrt{M_{\alpha/2}} X_{0,\Sigma}, \quad (2.8)$$

where M is distributed as above, and $X \sim N_d(0, \Sigma)$.

Thus, it is also very easy to simulate from the multivariate Linnik distribution, since it is simple to draw random samples from a multivariate normal distribution. This gives rise to the following simple algorithm for sampling from the multivariate Linnik distribution:

- (a) Draw $m \sim \text{Mittag - Leffler}(\alpha/2)$.
- (b) Draw $\mathbf{x} \sim N(0, \Sigma)$ independently. Set $\mathbf{y} = \sqrt{m}\mathbf{x}$.

3. A Multivariate Linnik Copula

In this section a novel elliptical Linnik copula is presented. The first subsection focuses on the derivation of this copula through the fact that a Linnik distribution is part of the larger family of elliptical distributions. Next, two methods of drawing samples from this copula are described. Finally, a numerical estimation procedure is presented.

3.1 Elliptical Linnik Copulas

A copula can be constructed from the multivariate Linnik distribution using the knowledge that a multivariate Linnik distribution is elliptical. Thus, the Linnik distribution fits the definition of an elliptical distribution given in Section 3. Additionally, using Eqn (2.25) we can see that the relevant characteristic function, φ , is of the form,

$$\varphi(\mathbf{t}'\Sigma\mathbf{t}) = \frac{1}{1 + (\mathbf{t}'\Sigma\mathbf{t})^{\alpha/2}}.$$

Then, a Linnik copula follows from the definition of elliptical copulas,

$$C(x_1, \dots, x_d) = F_{\alpha,\Sigma}(F_1^{-1}(x_1), \dots, F_d^{-1}(x_d)), \quad (3.1)$$

where $F_{\alpha,\Sigma}$ is the multivariate distribution function of the Linnik distribution and $F_1^{-1}, \dots, F_d^{-1}$ are the quantile functions of the Linnik distribution with unit spread. Unfortunately, no analytic form can be easily found with this construction. Even a copula density function is difficult to find analytically and may require using numerical differentiation.

3.2 Simulation

In this subsection, two methods of simulating data using the above defined copula are presented. One utilizes the Linnik distribution function, while the other bypasses this to utilize the empirical distribution function. They are demonstrated to be practically equivalent.

The first approach uses the inverse transform of Sklar's Theorem (in the bivariate case) defined as

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), \quad (3.2)$$

where the marginal distributions of x_i are $\sim L(\alpha, \gamma_i)$, where α is the characteristic exponent of the distribution, and γ_i are the diagonal entries of the covariance matrix Σ .

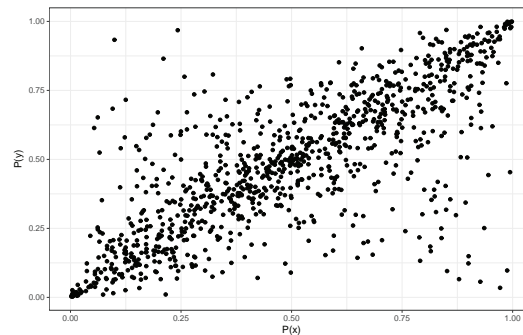
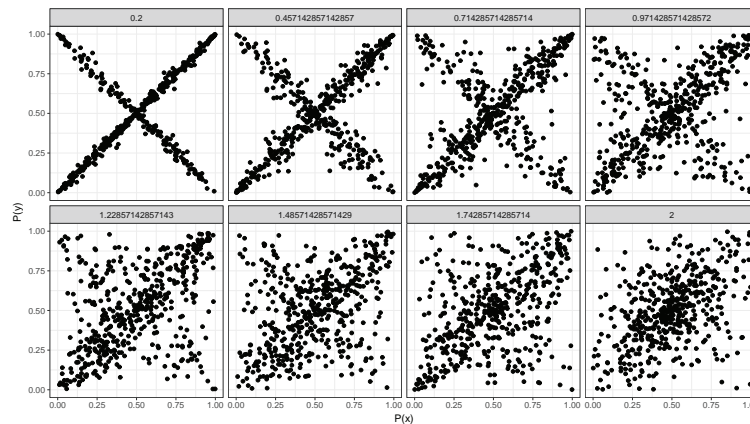


Figure 3.1. Scatterplot of 1000 samples of the bivariate Linnik copula

Figure 3.2. Scatterplots of 1000 samples from the Linnik copula for 8 different values of α with constant $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{21} = \Sigma_{12} = 0.5$.

This approach gives rise to a relatively straightforward simulation algorithm to draw samples from a Linnik copula (with any marginals). First, draw samples from the multivariate Linnik copula (using the Algorithm from Section 3.3). Then, numerically evaluate the density function of the univariate Linnik distribution for these samples. Finally, one only needs to use the quantiles of the marginal distributions to obtain samples from the copula based distribution. Of course, this method assumes that the quantiles are known for the marginal distributions, but in practice, this is a low barrier. This method is of course equivalent to the standard method for drawing samples from the elliptical copula, given in Example 3.2.3.

An example of the scatterplot of 1000 samples from the Linnik copula with parameters $\alpha = 1.4$ and $\Sigma = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$ is shown in Fig. 4.1.

Several scatterplots of 1000 samples for different values of α while $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ are shown in Fig. 4.2, while in Fig. 4.3 $\alpha = 1.4$, and Σ varies.

One issue with the sampling method described above is that the probability estimation part is quite slow and potentially inaccurate. For example, producing the scatterplot of just 1000 samples from a two-dimensional Linnik distribution takes about 29 seconds.

An alternative method is to use approximate probabilities for each of the samples in the distribution and to use the empirical probabilities generated using the empirical cdf. This is a very similar approach to the empirical copula function, and could be understood as estimating an empirical copula from samples of the Linnik distribution, but then using the copula to draw samples from another distribution.

In practice, this method produces very similar results to the actual Linnik copula, but is much quicker. An experiment with 10 replications of the Kullback-Leibler (KL) distance between samples of increasing size from the empirical cdf copula and the exact copula was performed, with results shown in Fig. 4.4.

The results of this experiment show the quick convergence of samples from the empirical copula to the Linnik copula for sample sizes of at least 325. This suggests that when the Linnik copula is modeled on real world data sets, using the

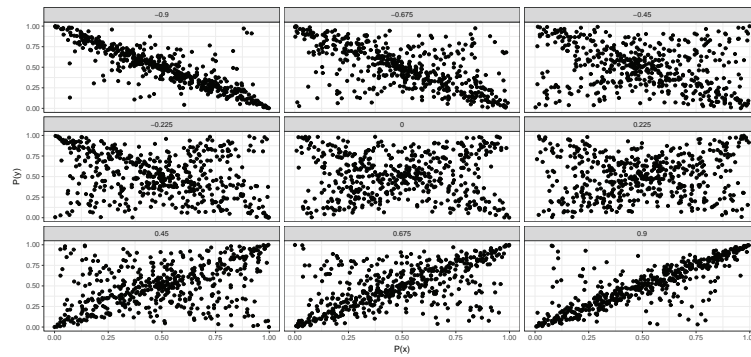


Figure 3.3. Scatterplots of 1000 samples from the elliptical Linnik copula for 9 different values of Σ , with $\alpha = 1.4$

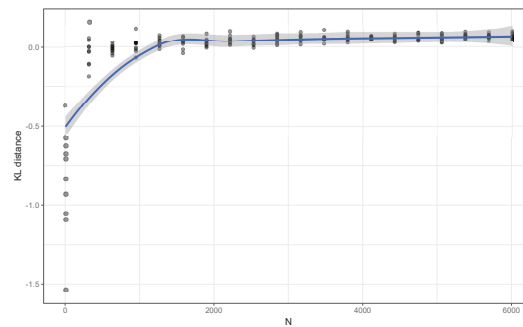


Figure 3.4. Plot of KL distance between samples using empirical copula and exact copula. Blue line shows the smoothed distance estimate, while gray shading represents the 95% confidence band.

empirical copula is an effective, speedy and highly accurate alternative. This convergence result is due to the convergence properties of the empirical copula to the true copula function (see Eqn. 2.7).

4. Dependency Measures for Linnik Random Variables

4.1 General Properties of Dependency Measures

Quantifying the dependence between random variables is a fundamental problem in statistics and many related issues remain unsolved. Metrics that are commonly used often pick up a limited class of associations, are computationally inefficient, or suffer from a lack of interpretability. In this section, we describe the requirements for the dependency measures, review several popular measures of statistical dependence and present results of experiments estimating dependency measures for Linnik copulas.

In (1959) classic paper on dependency Alfred Renyi identified seven properties that a measure of association, $\delta_{X,Y}$, must satisfy for any two random variables X and Y :

1. $\delta_{X,Y}$ is defined for every pair X and Y .
2. Symmetry: $\delta_{Y,X} = \delta_{X,Y}$.
3. Range: $0 \leq \delta_{X,Y} \leq 1$.
4. Independence: $\delta_{X,Y} = 0$ if and only if X and Y are independent.
5. Monotonicity: $\delta_{X,Y} = 1$ if and only if X and Y are almost surely strictly monotone functions of each other.
6. Invariance: if f and g are almost surely strictly monotone functions on the range of X , and range of Y , respectively (i.e., bijective Borel-measurable functions), then $\delta_{f(X),g(Y)} = \delta_{X,Y}$.
7. Pointwise convergence: if X_n, Y_n are sequences of random variables with joint distribution H_n that converges pointwise to the joint distribution H , then $\lim_{n \rightarrow \infty} \delta_{X_n, Y_n} = \delta_{X,Y}$.

Several variants of this definition exist (see, e.g., Nelsen (2006), and Lopez-Paz (2013)), but the key details in all those definitions are the same.

4.2 Useful Dependency Metrics

In this subsubsection, several different dependency metrics that account for various types of dependence are reviewed. When possible, the type of statistical dependence that they account for is also given. For further detail on the first three metrics and their relation to copulas, see Chapter 5 of Nelsen (2006).

4.2.1 Pearson's Correlation

Pearson's correlation is a measure of the linear dependence between two random variables X and Y . We can recall from basic statistics that Pearson's correlation is a moment based measure, defined in the theoretical case as

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E(X)^2} \sqrt{E(Y^2) - E(Y)^2}}.$$

The existence of the second moments is a necessary condition for the existence of the correlation metrics. However, for all finite samples, the empirical correlation

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

is well defined.¹

4.2.2 Kendall's Tau

Similar to the preceding metric, Kendall's tau is another basic measure of dependence between two random variables. Defined simply, it measures the association between "large" values of a random variable X and "large" values of a random variables Y - that is the linear association between the ranks of the variables. Thus, in the sample version, it is defined as

$$\tau = \frac{c - d}{n(n-1)},$$

where c is the number of concordant pairs in the sample, ie those such that $x_i < x_j$ and $y_i < y_j$, and d is the number of discordant pairs in the sample, which is simply the number of pairs in the sample that are not concordant.

In the case when the copula function, $C(u, v)$, between the random variables is known, then an explicit population version can be derived:

$$\tau_{X,Y} = 4 \int \int_{\mathcal{I}^2} C(u, v) dC(u, v) - 1.$$

4.2.3 Spearman's Rho

Spearman's rho rank correlation coefficient is a third basic measure of linear dependence between two random variables. Similar to Kendall's tau, it operates over the ranks of the random variables, but is defined like the standard Pearson's correlation coefficient, just over the ranked (ordered) random variables, instead of the random variables themselves. From a random sample, it can be defined as $S\rho = \frac{\text{Cov}(\text{rank}(x), \text{rank}(y))}{\sigma_{\text{rank}(x)} \sigma_{\text{rank}(y)}} = 1 - \frac{6 \sum_i \text{rank}(x_i) - \text{rank}(y_i)}{n(n^2-1)}$,

where $\sigma_{\text{rank}(x)}$, and $\sigma_{\text{rank}(y)}$, are standard deviations of the rank variables/ The second equality is valid only if all ranks are distinct integers.

Similar to Kendall's tau, if the copula function describing two random variables is known, then we have the following expression for the Spearman's rho correlation coefficient :

$$\rho_{X,Y} = 12 \int \int_{\mathcal{I}^2} uv dC(u, v) - 3.$$

This expression gives rise to bounds on the following relation between Spearman's rho and Kendall's tau, (see also Nelsen (2006))

$$-1 \leq 3\tau - 2\rho \leq 1.$$

4.2.4 Szekely's Energy Correlation

Szekely's distance correlation is an energy-based measure of correlation that tests the for independence of the characteristic functions of two random variables (see Szekely, Rizzo and Bakirov (2007)). Simply put, it is a hypothesis test against

¹The standard R implementation was used for all work in this thesis.

the null hypothesis that the joint characteristic function and the product of the two characteristic functions are equal to each other. Under the assumption that both distributions on X and Y have finite first moments, the distance correlation between two random variables (or potentially vectors, X and Y) is given by the formula,

$$dCor(X, Y) = \begin{cases} \frac{dCov(X, Y)}{\sqrt{dVar(X)dVar(Y)}}, & \text{if } dVar(X)^2 dVar(Y)^2 > 0 \\ 0, & \text{else} \end{cases}, \quad (4.1)$$

where $dCov(X, Y)$ is the distance covariance and can be represented in terms of characteristic functions as

$$dCov(X, Y)^2 = \frac{1}{c_p c_q} \int_{-\infty}^{\infty} \frac{|\phi_{X,Y}(s, t) - \phi_X(s)\phi_Y(t)|^2}{s^2 t^2} ds dt, \quad (4.2)$$

where $c_p = c_q = \frac{\pi}{\Gamma(1)} = \pi$ if both distributions are one-dimensional. Similarly, $dVar(X) = dCov(X, X)$.

One can define a variant of the distance correlation that requires only existence of the moments of fractional order $\beta \in (0, 2)$. This can be defined explicitly in the multivariate case, however, in the bivariate case, the distance covariance is defined as

$$V^2(X, Y) = \frac{1}{\pi^2} \int_0^{\infty} \int_0^{\infty} \frac{|\phi_{X,Y}(s, t) - \phi_X(s)\phi_Y(t)|^2}{(st)^{\beta+1}} ds dt. \quad (4.3)$$

Now, only the β -th moment needs to exist in order for the distance covariance to be finite. Distance correlation can then be constructed from the distance covariance and variance in exactly the same way as Equation 4.2. Unsurprisingly, this metric can be numerically found with a similar algorithm as the standard distance correlation. One substitution must be made however, in that the β th moment must be substituted for the first moment in the integral.

4.2.5 Schweizer and Wolff's Sigma

Schweizer and Wolff's Sigma is another copula based method of quantifying the dependence. It is defined by the formula (see Nelsen (2006)),

$$\sigma = 12 \int_{I^2} |C(u, v) - uv| du dv.$$

and geometrically, this can be interpreted as taking the L_1 - norm of the distance between the specified copula function and the independence copula function. Unfortunately, this expression requires explicit knowledge of the copula that describes the data. However, when the copula is not known the method can still be used by an application of the empirical copula function.

4.2.6 Randomized Dependence Coefficient and Tail Dependence

The Randomized Dependence Coefficient (RDC) is a relatively new method for determining the dependence between two random variables, first published in the 2013 paper by Lopez-Paz, Hennig and Schoelkopf. It is described as measuring the "largest canonical correlation between k randomly chosen non-linear projections of their [empirical] copula transformations". This coefficient comes out of the machine learning literature where it has been successfully used for feature selection in high dimensional random variables. For details see the above cited paper.

The tail dependence of a bivariate distribution on X, Y with distribution functions F, G is a measure of how dependent the tails of the distribution are on each other (see Nelsen (2006)). The lower tail dependence is the limit of the (conditional) probability that Y is less than the $100u$ -th percentile of G given that X is less than or equal to the $100u$ -th percentile of F , that is

$$\lambda_L = \lim_{u \rightarrow 0} P(Y \leq G^{-1}(u) | X \leq F^{-1}(u)).$$

Similarly, the upper tail dependence is the limit of the conditional probability that Y is greater than the $100u$ -th percentile of G given that X is greater than or equal to the $100u$ -th percentile of F , that is

$$\lambda_U = \lim_{u \rightarrow 1} P(Y > G^{-1}(u) | X \leq F^{-1}(u)).$$

This has a copula based definition, where

$$\lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} = c(0, 0)$$

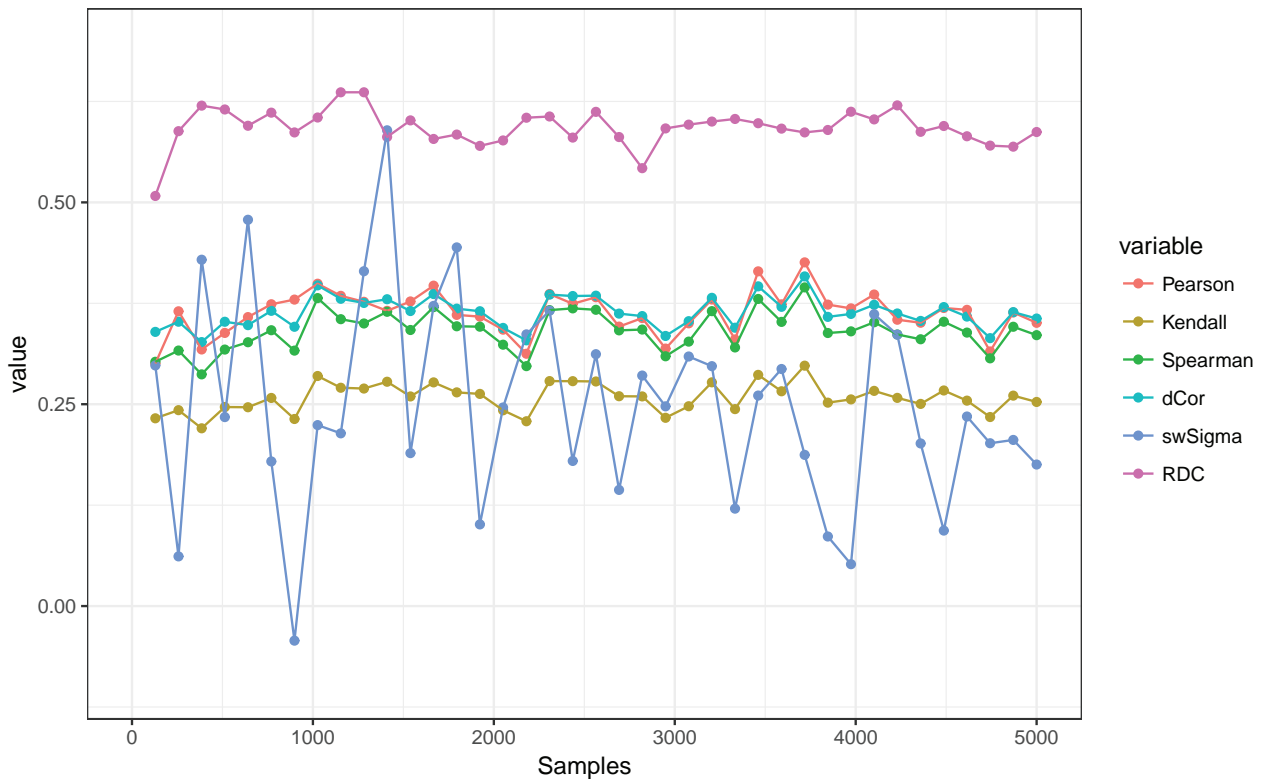


Figure 4.1. Plots of different numerically computed dependency metrics for Linnik copula as functions of the number of samples.

and

$$\lambda_U = 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} = 2 - c(1, 1).$$

Estimates of these coefficients could potentially be evaluated using the empirical copula. Unfortunately, these estimates may not perform well in practice; see Frahm, Junker, and Schmidt (2005) for a nice discussion of the difficulties in estimating the tail dependence from samples of bivariate distributions.

4.2.7 Numerical Experiments Evaluating Dependency Metrics

To test the effectiveness of the above dependency metrics in capturing the type of association in the Linnik copula, we conducted numerical experiments in which 40 samples of increasing sample size (from 2 to 5000) were drawn from a Linnik copula with $\alpha = 1.4$ and $\Sigma = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$. Then, the dependence metrics described in subsections 5.2.1–5.2.6 were evaluated over these samples. The purpose of the test was twofold; first, to see which metrics (if any) would converge over the increasing sample size, and second, to see what values of different dependence metrics would be produced. The results are shown in Fig. 5.1.

Interestingly, the randomized dependence coefficient RDC showed the highest value of dependence. This may be a property of the definition of this coefficient. Experiments in Lopez-Paz, Henning and Scheolkopf (2013) showed that the RDC gave estimates similar to Pearson's rho when the relationship is Gaussian. Perhaps the RDC acts in an additive manner, increasing the value when the relationship is more than Gaussian.

Pearson's correlation, distance correlation and Spearman's rho all quickly reached values of about 0.33. This may be due to the value of Σ_{12} in the experiment, but is currently unknown. Kendall's tau sat around 0.25 for the entire experiment, and it is again unknown what the meaning of this value is. Finally, Schweizer and Wolff's sigma did not converge over the entirety of the experiment for unknown reasons. Overall, RDC seems to outperform the other metrics in this experiments.

5. Estimation of Parameters for Multivariate Linnik Distributions

We have tested several estimation approaches including the empirical characteristic functions and the method of logarithmic moments which although efficient in the one-dimensional case (see Kozubowski (2001), and Cahoy (2012, 2013)) did not perform well for the multivariate case. The only approach which displayed good convergence to theoretical values of

the multivariate parameters involved was the divergence minimization method which is described below.

5.1 Multivariate Parameter Estimation Using Divergence Minimization

Approaching parametric estimation as divergence minimization has a long history that is explored in greater detail in Basu, Shioya, and Park (2011). With this method, parametric estimation is done through the numerical minimization of a given divergence between the population data and simulated data (with the given parameters). That is, to produce estimates of the parameters, we only need to solve the following minimization problem,

$$\arg \min_{\theta} d(p(x), q(x; \theta)),$$

where $p(x)$ is the true population distribution function and $q(x; \theta)$ is the distribution function with parameters θ . In practice, we do not know the population distribution, but have only samples from it. Thus, we must use a discrete approximation of the divergence function $d(., .)$ and samples from the distribution $q(x; \theta)$.

For estimation of the parameters of the multivariate Linnik distribution, we can easily get a large sample with given parameters, and as an easily computable divergence function we selected is the Kullback-Leibler (KL) divergence (see, Kullback, Leibler (1951)), defined as

$$D_{KL}(p||q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx,$$

where $p(x)$ and $q(x)$ are density functions that may or may not be vector valued (see Joyce (2011)). This divergence function is widely used across most mathematical literature due to its nice properties. In particular, a maximum likelihood estimator will minimize the KL divergence between the distributions, and directly minimizing the KL divergence will produce an estimator identical with the maximum likelihood estimator (see Joyce (2011), and Basu, Shioya, and Park (2011)). A useful numerical implementation of the KL divergence in high dimensions is in the FNN R package² which uses a k-nearest neighbors approach to compute the divergence in high dimensions (see Boltz, Debreuve and Barlaud (2007)).

Thus, the minimization problem to be solved consists of simply finding ³

$$\arg \min_{\alpha, \Sigma} KL(p(x)||q(x; \alpha, \Sigma)). \quad (5.1)$$

Here, the equivalence with the maximum likelihood estimator can be clearly seen. Recalling the definition of the KL divergence, we can write it as

$$KL(p(x)||q(x; \alpha, \Sigma)) = \int_{-\infty}^{\infty} [p(x) \log(p(x)) - p(x) \log(q(x; \alpha, \Sigma))] dx.$$

Separating out the logarithm gives

$$KL(p(x)||q(x; \alpha, \Sigma)) = E_{p(x)}(\log(p(x))) - E_{p(x)}(\log(q(x; \alpha, \Sigma))).$$

Given a dataset $p(x)$, $E_{p(x)}(\log p(x))$ will be constant, so minimizing the KL divergence is just maximizing $E_{p(x)}(\log q(x))$. If the distributions p and q are the same, this is just maximizing the log-likelihood of the data, which corresponds to the maximum likelihood estimation.

One popular algorithm for optimization, particularly in a noisy setting, is the simultaneous perturbation stochastic approximation (SPSA) algorithm invented by James Spall (1998). SPSA is an iterative procedure where the parameter value changed via a subtraction of the gradient at each step. However, the gradient is calculated through a simultaneous perturbation of the current theta value, which reduces the number of loss function calculations to two at each step (versus $2p$ for traditional gradient descent). Ultimately, this vastly reduces the computational time of the algorithm.

5.2 Test of Numerical Estimation Procedure

Starting values of the parameters for the SPSA algorithm were chosen as follows: $\alpha = 0.8$, $\Sigma_{12} = 0.4$, and $\sigma_1^2 = \sigma_2^2 = 1$. Due to the multi-dimensional optimization, and the repeated KL-divergence calculations, this procedure is not instantaneous, and takes a fairly long time to run. A plot of the time for each iteration of the experiment is shown in Fig. 6.1. The estimated parameter values for each iteration are shown in Fig. 6.2. Finally, the KL distance was recalculated and is shown in Fig. 6.3.

With a few exceptions, the time of estimation grows linearly with the sample size, ranging from about 1 second for a sample size of 10 to about 250 seconds for the largest sample size, of 50,000.

²<https://CRAN.R-project.org/package=FNN>

³The KL divergence is anti-symmetric. That is, $KL(p||q) \neq KL(q||p)$.

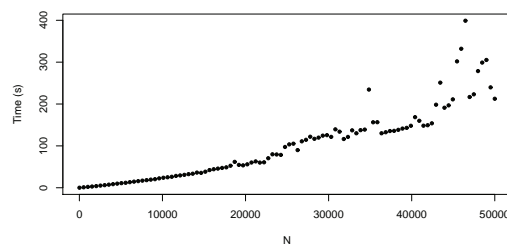


Figure 5.1. Plot of the time needed for each iteration of parameter estimation experiments using divergence minimization.

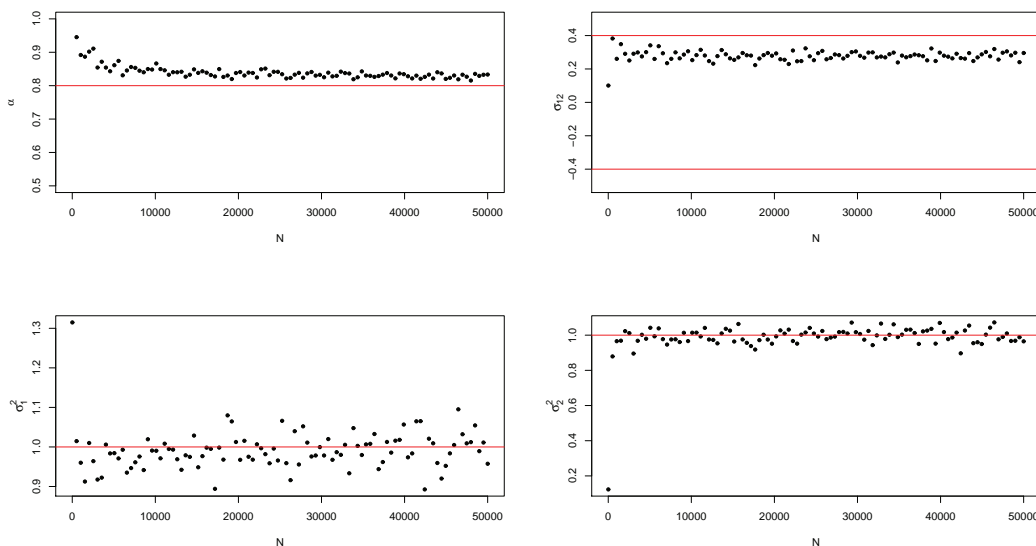


Figure 5.2. Estimated values for true parameters $\alpha = 0.8$ (top left), $\Sigma_{12} = 0.4$ (top right), and $\sigma_1^2 = \sigma_2^2 = 1$ (bottom) as functions of the sample size.

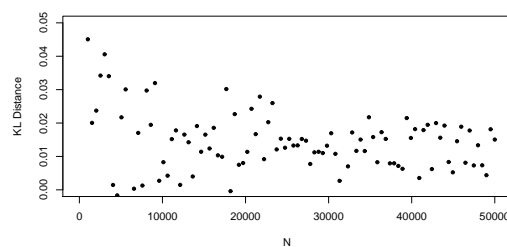


Figure 5.3. Plot of calculated KL distance for experiments using divergence minimization. Potentially negative values are due to error in numerical calculation.

Overall, the KL divergence minimization gives more accurate results than the MoM estimator, but comes at the price of speed. This is largely due to the fact that estimating the KL divergence requires a numerical integration. In practice, this speed trade-off should not be too terrible, but could potentially be an issue in high dimensions, or when updating the parameters over time is required.

References

Anderson, D. N. (1992). A multivariate Linnik distribution. *Statistics & Probability Letters*, 144, 333 - 336.

[https://doi.org/10.1016/0167-7152\(92\)90067-F](https://doi.org/10.1016/0167-7152(92)90067-F)

- Basu, A., Shiyoa, H., & Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. CRC Press, Chapman and Hall.
- Bertoin, J. (1996). *Lévy Processes: Theory and Applications*. Cambridge University Press. Cambridge.
- Boltz, S., Debreuve, E., & Barlaud, M. (2007). kNN-based high-dimensional Kullback-Leibler distance for tracking, In *Image Analysis for Multimedia Interactive Services* WIAMIS International Workshop, 16pp., IEEE.
- Cahoy, D. O. (2012). An estimation procedure for the Linnik distribution. *Statistical Papers*, 53(3), 617 - 628. <https://doi.org/10.1007/s00362-011-0367-4>
- Cahoy, D. O. (2013). Estimation of Mittag-Leffler Parameters. *Communications in Statistics - Simulation and Computation*, 42(2), 303C-315. <https://doi.org/10.1080/03610918.2011.640094>
- Chambers, J. M., Mallows, C. L., & Stuck, B. W. (1976). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association*, 71(354), 340. <https://doi.org/10.1080/01621459.1976.10480344>
- Devroye, L. (1990). A note on Linnik's distributions. *Statistics & Probability Letters*, 9, 305-306. [https://doi.org/10.1016/0167-7152\(90\)90136-U](https://doi.org/10.1016/0167-7152(90)90136-U)
- Frahm, G., Junker, M., & Schmidt, R. (2005). Estimating the tail-dependence coecient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37, 80-100. <https://doi.org/10.1016/j.insmatheco.2005.05.008>
- Gunaratnam, B., & Woyczynski, W. A. (2015). Multiscale Conservation Laws Driven by Levy Stable and Linnik Diffusions: Asymptotics, Shock Creation, Preservation and Dissolution, *Journal of Statistical Physics*, 160(1), 29 - 59. <https://doi.org/10.1007/s10955-015-1240-y>
- Joyce, J. M. (2011). Kullback-Leibler Divergence, *International Encyclopedia of Statistical Science*. M. Lovric, editor, Springer Berlin Heidelberg, 720C722.
- Kotz, S., Kozubowski, T. J., & Hayfavi, A. (1995). Analytic and asymptotic properties of Linnik's probability densities, *J. Math. Anal. Appl*, 193, 353-371 and 497-521.
- Kotz, S., Kozubowski, T. J., & Podgórski, K. (2001). *The Laplace Distribution and Generalizations*, Springer, New York. <https://doi.org/10.1007/978-1-4612-0173-1>
- Kozubowski, T. J. (1994). The inner characterization of geometric stable laws. *Statist. Decisions*, 12, 307-321. <https://doi.org/10.1524/strm.1994.12.3.307>
- Kozubowski, T. J. (2000). Computer simulation of geometric stable distributions, *Journal of Computational and Applied Mathematics*, 116, 221-229. [https://doi.org/10.1016/S0377-0427\(99\)00318-0](https://doi.org/10.1016/S0377-0427(99)00318-0)
- Kozubowski, T. J. (2001). Fractional moment estimation of Linnik and Mittag-Leffler parameters, *Mathematical and Computer Modeling*, 34, 353-371. [https://doi.org/10.1016/S0895-7177\(01\)00115-7](https://doi.org/10.1016/S0895-7177(01)00115-7)
- Kozubowski, T. J., & Panorska, A. K. (1999). Simulation of Geometric Stable and Other Limiting Multivariate Distributions Arising in Random Summation Scheme, *Mathematical and Computer Modeling*, 29, 252 - 262. [https://doi.org/10.1016/S0895-7177\(99\)00108-9](https://doi.org/10.1016/S0895-7177(99)00108-9)
- Kozubowski, T. J., & Rachev, S. T. (1999). Univariate geometric stable laws, *Journal of Computational Analysis and Applications*, 1(2), 177 - 217. <https://doi.org/10.1023/A:1022629726024>
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency, *Annals of Mathematical Statistics*, 22(1), 79 - 86. <https://doi.org/10.1214/aoms/1177729694>
- Li, D. (2000). On default correlation, *Journal of Fixed Income*, 9(4), 41-90. <https://doi.org/10.3905/jfi.2000.319253>
- Linnik, Y. V., Linear forms and statistical criteria. I, II, (1953). *Selected Transl. in Math. Statist. and Prob*, 3, 41 - 90.
- Lopez-Paz, D., Hennig, P., & Schoelkopf B. (2013). The randomized dependence coefficient. *Advances in Neural Information Processing Systems*, 1C9.
- Mai, J.-F., & Scherer, M. (2012). Simulating copulas: stochastic models, sampling algorithms, and applications. *Series in Quantitative Finance*, Vol 4, Imperial College Press ; World Scientific, London : Hackensack, NJ. <https://doi.org/10.1142/p842>
- Nelsen, R. B. (2006). *An introduction to Copulas*, Springer Series in Statistics. Springer, New York, 2nd edition.
- Ostrovskii, I. (1995). Analytic and asymptotic properties of multivariate Linnik distributions, *Math. Phys. Anal. Geome-*

try, 2(3), 436-455.

- Póczos, B., Ghahramani, Z., & Schneider, J. (2012). Copula-based kernel dependency measures. In *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh.
- Renyi, A. (1959). On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae*, 10, 441-451. <https://doi.org/10.1007/BF02024507>
- Rey, M. (2015). Copula models in machine learning, *Ph.D. Thesis*, University of Basel.
- Ruschendorf, L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process, *Journal of Statistical Planning and Inference*, 139(11), 3921-3927. <https://doi.org/10.1016/j.jspi.2009.05.030>
- Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall.
- Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.
- Shephard, N. (1991). Numerical Integration Rules for Multivariate Inversion. *J. Statist. Comput. Simul*, 39, 37 - 46. <https://doi.org/10.1080/00949659108811337>
- Spall, J. C. (1998). Implementation of the simultaneous perturbation algorithm for stochastic optimization, *IEEE Transactions on Aerospace and Electronic Systems*, 34(3), 817-823. <https://doi.org/10.1109/7.705889>
- Szekely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances, *Annals of Statistics*, 35(6), 2769-2794. <https://doi.org/10.1214/0090536070000000505>
- Wilson, A., & Ghahramani, Z. (2010). Copula processes, *Advances in Neural Information Processing Systems*, 2460-2468.
- Zezula, I. (2009). On multivariate Gaussian copulas, *Journal of Statistical Planning and Inference*, 139(11), 3942-3946. <https://doi.org/10.1016/j.jspi.2009.05.039>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).