Multivariate measures of dependence for random variables and Lévy processes

Alexandru C. Belu

Submitted in partial fullfillment of the requirements for the degree of Doctor of Philosophy

Dissertation Advisor: Dr. Wojbor A. Woyczynski

Department of Statistics
Case Western Reserve University

May, 2012

CASE WESTERN RESERVE UNIVERSITY
SCHOOL OF GRADUATE STUDEIS

We hereby approve the dissertation of

ALEXANDRU C. BELU

candidate for the Doctor of Philosophy degree *

Committee Chair: ______________________________

Dr. Wojbor A. Woyczynski

Dissertation Advisor

Professor

Department of Statistics

Committee: ______________________________

Dr. Harold S. Haller

Adjunct Professor

Department of Statistics

Committee: ______________________________

Dr. Catalin Turc

Asisstant Professor

Department of Mathematics

Committee: ______________________________

Dr. Patricia Williamson

Instructor

Department of Statistics

March 13th 2012

*We also certify that written approval has been obtained for any proprietary material contained therein.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my family for their support both moral and financial troughout the five years of my dissertation. My parents, Radian and Paulina Belu, for always pushing me to achieve my full potential, while always being there to love me, pick me up and dust me off when things didn't go as planned. My sister, Maria Belu, for the herculean dedication to spell checking the countless pages of dry mathematics of my dissertation. My wife, Mirela Belu, for constantly asking me why I'm not done yet, and providing that extra impetus to keep on working, even when the going got tough. I would also like to thank Sharon Dingess, for her invaluable help navigating the adminsitrative pitfalls in my way. Last, but certainly not least, I would like to thank my advisor, Dr. Wojbor Woyczynski, from the bottom of my heart, for being a mentor and a friend for the duration of my PhD. I couldn't have asked for a better guide these past years and words cannot express how grateful I am for being lucky enough to have studied under him.

List of Notations

$U[0,1]$	Continuous Uniform Distribution on [0,1]
Ran X	Range of random variable X
Dom X	Domain of random variable X
Ran F	Range of function F
Dom F	Domain of function F
c.d.f., CDF, and d.f.	cumulative distribution function
i.i.d	independent and identically distributed
ID	infinitely divisible
$\bar{\mathbf{R}}$	$\mathbf{R} \cup \{-\infty, \infty\}$
$\mathbf{R}_+$	The set of non-negative reals
Π	The independence copula
M	The maximum copula
W	The minimum copula
c.f.	characteristic function
F.G.M.	Farlie-Gumble-Morgenster distribution
a.s.	almost surely
$< a, b >$	scalar product of vectors a and b
$\mathcal{L}$	Likelihood function
rcll	right continuous with left limits (a.k.a. cadlag)
$Y \stackrel{d}{=} X$	Y is equal to X in distribution
rv	random variable
$\mathbf{Z}$	The set of integers

Multivariate Measures of Dependence for Random Variables and Lévy Processes

Abstract

by

ALEXANDRU C. BELU

Multipoint correlations and measures of association that are applicable to random variables lacking first or higher moments are becoming increasingly important. Increasing computational power as well as the proliferation of massive data, has spurred research into multipoint correlations in an attempt to measure the association between several random variables at once. At the same time, advances in stochastic processes and Lévy tempered stable distributions, has introduced the need for measures of dependence which can be applied to volatile distributions lacking first or second moments. In this dissertation, we present two such measures for studying the dependence of distributions without higher moments and multipoint correlations. We present a thorough overview of copulas, Lévy processes and Lévy copulas, as well as introduce the two measures in question, Schweizer and Wolffs sigma and Szekelys distance correlation. We extend these measures to the multivariate case, and provide sample estimates for each. We take these measures and apply them to various bivariate distributions with specific properties and examine their behavior in these cases. Finally we introduce our very own measure of dependence for Lévy processes, setting the groundwork for establishing a measure of association between such processes.

Keywords: copulas, Lévy copulas, Lévy processes, measures of dependence, Szekely's distance correlation, Schweizerr and Wolff sigma, generating bivariate data, bivariate distributions, multivariate measures of association, multipoint correlations.

1 Introduction

Analyzing the correlation or dependence between random variables is a rich field comprised of many different measures and ways to look at the problem. From the classical Pearson's Product Moment Correlation, to Spearman's Rank Rho, Measures of Concordance and more exotic methods based on characteristic functions or Shannon's Information Theory, the literature is diverse and extensive. Applying these techniques to α-stable distributions and Lévy Processes poses a whole new set of challenges, since most of these distributions do not have second or sometimes first moments, a requirement for measures such as Pearson's Product Moment Correlation.

In this thesis we embark on a comprehensive study of existing measures which relax these moment requirements, and in some cases extend these to the multivariate case, when such an extension is found lacking. The extension to the multivariate case serves as a preamble into the growing field of multipoint correlations. As the availability of massive data sets with a large number of parameters becomes more widespread throughout all scientific fields, we see a growing interest in the ability to measure correlation between three or more random variables.

Examples of work which studies multipoint correlations abounds across the many scientific disciplines. Šanda and Mukamel studied in relations to anomalous diffusion processes in chemistry [32]. Miller, Dasi and Webster analyzed concentration fluctuations in turbulent fields using multipoint correlations with applications to fluid dynamics [17]. Similarly Dmowski used multipoint correlations to study deep-level measurements in metal-oxide-semiconductor devices [6], and Niemann and Kantz used the concept of multipoint correlation and applied it to continuous-time random walks [20].

During our work we found two primary such measures, Szekely's Dependence, and the Schweizer and Wolff's Sigma, which we analyzed. We looked at the relationship between them in the context of specific bivariate distribution functions and bivariate families of distributions, both through study of analytic expressions as well as computer simulation and random sample estimates.

In closing we developed and refined our own measure of dependence for Lévy processes based on Lévy copulas, and we check that it is indeed a measure of dependence based on the updated Rényi dependence properties.

1.1 Topics Covered

We begin in **Chapter 1**, entitled **Mathematical Background**, by embarking on a thorough understanding of the definitions and concepts that introduce the general concept of a Dependence Measure. We start by presenting the principal properties of a dependence measure as well as trying to understand the main concepts they entail.

We also present some of the small differences in definitions we encounter across the myriad of papers and books on the subject, and emphasize the important basic structure that they are all built upon. The reader is also treated to a thorough introduction to copula and Lévy copula concepts, as they are both at the center of this thesis, and knowledge of relevant theorems is necessary for a thorough understanding of the thesis.

In **Chapter 2**, the two most important and most useful dependence measures we have found are given a thorough presentation, including principal theorems and proofs that go along with them. The first of these is **Szekely's Distance Correlation**, a measure of dependence based on characteristic functions, and one which shows great promise for dealing with distributions and stochastic processes lacking second or even first moments and devoid of a clearly defined distribution and density function, a prerequisite of the other two dependence measures. The second is **Schweizer and Wolff's Sigma**, a measure of dependence, that takes advantage of copulas and their growing popularity, and gives us the insight needed to produce our own measure of dependence for Lévy processes, based on Lévy copulas.

We continue in **Chapter 3** by presenting **Applications to Specific Bivariate Distributions** for both of these measures as well as analyzing the relationship between them within these cases. The primary distributions analyzed are the Bivariate Normal, Farlie Gumble Morgenster family of bivariate distributions, as well as the D.J. Marcus example, an interesting case of a bivariate distribution with stable marginals, that is not itself stable.

We continue in **Chapter 4** by introducing a **A Dependence Measure for Lévy Processes**. This measure is first defined for Lévy processes with positive jumps, and then extended to general Lévy processes.

We conclude the thesis in **Chapter 5** with a summary of the topics presented as well as future work extending our Lévy processes dependence measure to the multipoint case, and providing empirical equations for applications to wind speed samples.

In **Appendix A**, the reader is treated to an **Overview of Algorithms and Code** as well as a brief look, in **Appendix B**, at **Other Metrics**. Since we attempt to show how each of the dependence measures and their relevant extensions can be used in practice on a random sample, this thesis makes exensive use of mathematical software and programming langauges to generate random samples and perform our calculations. In this section we present these techniques as well as some of the pitfalls and problems encountered when trying to apply theory into practice. We also present a few other metrics which can help one understand the relationship between random variables beyond the usual measures of association.

2 Mathematical Background

In this section, we cover the essential mathematical background required to digest the rest of the thesis as well as a few extra bits of information we found exceedingly useful for anyone wishing to work with copulas or Levy copulas. While the work is not our own, the totality of the information contained within this chapter is found scatterd in various books and articles, and required an extensive effort to gather and systematically put together, thus producing the thorough reference source we had in mind for this chapter.

The first section discusses properties of measures of dependence, which were first introduced by Renyi [22] but were later modified to be more flexible and inclusive. The section finishes with an overview of copulas, Levy processes, and Levy copulas, giving the reader all the theorems and notions necessary to easily access the work in the later chapters.

2.1 Measures of Dependence

Measures of dependence are a particular way of looking at the association between random variables. What sets measures of dependence apart from other similar measures of association, is that one can accurately measure full independence between random variables as well as full dependence. The first attempt to describe the properties of measures of dependence is due to Rényi [22]. It turns out however, as shown even by Rényi himself, that these were far too restrictive and only the maximal correlation coefficient satisfied them all. Over the years these properties have been refined somewhat, and we present here the commonly accepted properties any given measure of dependence should satisfy.

Definition 2.1.1. A numeric measure δ of association between two continuous random variables X and Y is a *measure of dependence* if it satisfies the following:

1. *Well-Posedness:*

   $\delta_{X,Y}$ is defined for every pair of random variables X and Y;

2. *Symmetry:*

   $\delta_{X,Y} = \delta_{Y,X}$;

3. *Normalization:*

   $0 \leq \delta_{X,Y} \leq 1$;

4. *Extreme Cases:*

   (a) $\delta_{X,Y} = 0$ if and only if X and Y are independent;

   (b) $\delta_{X,Y} = 1$ if and only if each of X and Y is a.s. a strictly monotone function of the other;

5. *Invariance under monotone transformations*

   if f and g are almost surely strictly monotone functions on Ran X and Ran Y, respectively, then $\delta_{f(X),g(Y)} = \delta_{X,Y}$;

6. *Continuity:*

   if (X, Y) and $(X_n, Y_n), n = 1, 2, \ldots,$ are pairs of random variables with joint distributions H and H_n, respectively, and if the sequence $\{H_n\}$ converges weakly to H, then $\lim_{n\to\infty} \delta_{X_n,Y_n} = \delta_{X,Y}$.

2.2 Copulas

The term Copula, originally coined by Sklar in 1959 [29], comes from the Latin noun meaning ”a link, tie, bond”. A copula is defined as a function which links multivariate distribution functions to their one-dimensional marginals. Generally it is a multivariate distribution function defined on the unit n-cube $[0, 1]^n$, with uniformly distributed marginals. Mathematically, the argument that leads to copulas can be made by first noting that for two absolutely continuous random variables X_i with d.f.s F_i, for $, i = 1, 2, U_i = F_i(X_i)$ both are uniformly distributed random variables on $[0, 1]$. Hence for some joint d.f. F with marginals d.f.s F_1, F_2

$$\begin{aligned} F(x_1, x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) \\ &= P(F_1(X_1) \leq F_1(x_1), F_2(X_2) \leq F_2(x_2)) \\ &= P(U_1 \leq F_1(x_1), U_2 \leq F_2(x_2)) \\ &= C(F_1(x_1), F_2(x_2)). \end{aligned}$$

The function $C(x, y)$ above is a copula, a joint distribution function defined on the unit square $[0, 1]^2$ with standard uniform marginal distribution functions. It is worthwhile to note that there is no mathematical requirement to transform the marginals to uniform distribution functions on $[0, 1]$. In fact, Hoeffding used the interval $[-\frac{1}{2}, \frac{1}{2}]$ in his 1940 papers, and in Extreme Value Theory transforming to Frechet marginals is often the case. The transformation of the marginal distribution functions to $U[0, 1]$ has become an unspoken standard and does present one with an intuitive understanding of the concepts behind copulas.

Copulas are important to us not only because they provide a flexible tool for analyzing the association between random variables, without resorting to the more cumbersome joint distribution function, but they are also the foundation for one of the measures of dependence we will be studying in this thesis, namely the Schweizer and Wolff sigma.

2.2.1 Definitions and Theory

In what followes we present a collection of Theorems and Definitions that are central to a solid understanding of copulas. They have been collected from Nelson's book on copulas [18], but can be found in one form or another in any of the myriad of papers published on the subject in the past two decades.

Definition 2.2.1. Let $S_1, S_2, \ldots, S_n$ be nonempty subsets of the extended real line $[-\infty, \infty] = \bar{\mathbf{R}}$, and let H be a real valued function of n variables whose domain, Dom H, is given by Dom $H = S_1 \times S_2 \times \ldots \times S_n$. Let $[\mathbf{a}, \mathbf{b}]$ denote the n-box $B = [a_1, b_1] \times \ldots \times [a_n, b_n]$ whose vertices $\mathbf{c} = (c_1, \ldots, c_n)$ are in Dom H, and each c_k is either equal to a_k or b_k. Then the H-volume of B is given by

$$V_H(B) = \sum_{\mathbf{c} \in B} \operatorname{sign}(\mathbf{c}) H(\mathbf{c}),$$

where sign(c) is given by

$$\operatorname{sign}(\mathbf{c}) = \begin{cases} 1 & \text{if } c_k = a_k \text{ for an even number of k's;} \\ -1 & \text{if } c_k = a_k \text{ for an odd number of k's.} \end{cases}$$

Definition 2.2.2. An n-dimensional copula is a function $C : [0,1]^n \to [0,1]$, with the following properties:

1. C is *grounded*, that is for every $\mathbf{u} = (u_1, \ldots, u_n) \in [0,1]^n, C(\mathbf{u}) = 0$ if at least one coordinate $u_i = 0, i = 1, \ldots, n$.

2. C is *n-increasing*, that is for every $\mathbf{u} \in [0,1]^n$ and $\mathbf{v} \in [0,1]^n$ such that $\mathbf{u} \leq \mathbf{v}$, the C-volume $V_C([\mathbf{u}, \mathbf{v}])$ of the box $[\mathbf{u}, \mathbf{v}]$ is non-negative, where the C-volume $V_C([\mathbf{u}, \mathbf{v}])$ is defined as in definition 2.2.1 above.

3. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$, for all $u_i \in [0,1], i = 1, 2, \ldots, n$.

In the two-dimensional case the above simplifies to a more manageable defintion:

Definition 2.2.3. A bivariate copula is a function $C : [0,1]^2 \to [0,1]$, with the following properties:

1. C is *grounded*: $\forall u, v \in [0,1], C(u,0) = C(0,v) = 0$.

2. C is *2-increasing*: $\forall u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

3. $\forall u, v \in [0,1], C(u,1) = u$ and $C(1,v) = v$.

Sklar's theorem is perhaps the most essential result in the theory of copulas, and it is the foundation of many, if not most, of the applications of copulas throughout statistics.

Theorem 2.2.1. *(Sklar [28]). Let H be an n-dimensional distribution function with marginals $F_1, F_2, \ldots, F_n$. Then there exists an n-copula C such that for all $x_1, x_2, \ldots, x_n \in \bar{\boldsymbol{R}}$,*

$$H(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

Conversly, if C is an n-copula and $F_1, F_2, \ldots, F_n$ are distribution functions, then the function H is an n-dimensional distribution with marginals $F_1, F_2, \ldots, F_n$. Furthermore, if the marginals are all continuous, then C is unique. Otherwise, C is uniquely determined on Ran $F_1 \times$Ran $F_2 \times \ldots \times$Ran F_n, where Ran F_i is the range of the function F_i.

2.2.2 Properties

Along with the basic definitions and theorems, copulas also pose several properties that become very useful when trying to work and manipulate them and are listed below for the readers reference. For a more in depth look at copulas we recommend Nelson's book on the subject [18].

Theorem 2.2.2. *(Continuity[18] p.11). Let C be a bivariate copula. Then for all $u_1, u_2, v_1, v_2 \in [0,1]$ such tha $u_1 \leq u_2$ and $v_1 \leq v_2$,*

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|,$$

thus C is uniformly continuous on its domain.

Theorem 2.2.3. *(Differentiability[18] p.13). Let C be a bivariate copula. For any $v \in [0,1]$, the partial derivative $\frac{\partial C}{\partial u}(u,v)$ exists for almost all $u \in [0,1]$ and, for such v and u,*

$$0 \leq \frac{\partial C}{\partial u}(u,v) \leq 1.$$

A similar result holds for any $u \in [0,1]$. Furthermore, the functions $u \mapsto \frac{\partial C}{\partial v}(u,v)$ and $v \mapsto \frac{\partial C}{\partial u}(u,v)$ are well-defined and non-decreasing almost everywhere on $[0,1]$

Theorem 2.2.4. *(Invariance[18] p.25). Copulas are invariant under strictly monotone transformations of the random variables.*

Theorem 2.2.5. *(Frechet-Hoeffding Bounds[18] p.11). For every copula C, and every $(u,v) \in [0,1]^2$,*

$$\max(u+v-1, 0) \leq C(u,v) \leq \min(u,v),$$

where $W(u,v) = \max(u+v-1,0)$ and $M(u,v) = \min(u,v)$ are themselves copulas refered to as the minimum and maximum copulas.

2.2.3 Copulas and Association

Copulas are, at their core, an expression of the dependence between two or more random variables, and as such have been used to rewrite some of the more well known measures of association, as well as to produce new ones. Below we list the most important of these, the last being part of our principal focus on measures of dependence.

Kendall's Tau

Given random variables X and Y with joint distribution H, Kendall's tau is then defined as the difference between the probabilities of concordance and discordance for two independent copies (X_1, Y_1) and (X_2, Y_2) of (X, Y), that is

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Spearman's Rho

Given three independent copies $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) of a random vector (X, Y) with common joint distribution function H, Spearman's rho is defined as the difference between the probabilities of concordance and discordance of random pairs (X_1, Y_1) and (X_2, Y_3)

$$\rho = 3\left(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]\right).$$

Thus if we let C be the copula associated with (X, Y) distributed according to H, then we can rewrite these two measures of association as follows:

$$\tau = 4\int_0^1 \int_0^1 C(u,v)dC(u,v) - 1,$$

$$\rho = 12\int_0^1 \int_0^1 (C(u,v) - uv)dudv.$$

Schweizer and Wolff's Sigma

Like most measures of association, Kendall's Tau and Spearman's Rho suffer from the problem that they cannot be used to determine the independence of two random variables. To remedy this situation, we introduce our first measure of dependence based on copulas, which is usually called Schweizer and Wolff's Sigma. If we replace $(C(u,v)-uv)$ in the copula based definition of Spearman's Rho by its absolute value, we obtain our new measure of dependence σ, which is equal to zero if and only if the two given random variables are independent. More precisely,

$$\sigma = 12\int_0^1\int_0^1 |C(u,v)-uv|dudv.$$

2.2.4 Examples of Copulas

Below we present a few examples of the best known and widely used copulas, but this is by no means an exhaustive accounting of all possible copulas. They present a very good starting point for finding a copula to fit most needs. It is worthwhile to mention here that copulas strictly capture the dependence structure between random variables, and thus different random variables, with different marginal distributions can share the same copula if their dependence structure is the same. This highlights the versatility and usefulness of the copula function.

Example 1. *Marshal-Olkin family (1967)*
If $\alpha,\beta \in [0,1]$, then the function $C_{\alpha,\beta} : [0,1]^2 \to [0,1]$, defined by

$$C_{\alpha,\beta}(u,v) = \min(u^{1-\alpha}, uv^{1-\beta}),$$

is a bivariate copula.

Example 2. *Bivariate Pareto copula*
This copula is defined by the following formula

$$C_\alpha(u,v) = u+v-1+[(1-u)^{-\frac{1}{\alpha}}+(1-v)^{-\frac{1}{\alpha}}]^{-\alpha},$$

where α is a parameter ($\alpha \in R\backslash\{0\}$).

Example 3. *Farlie-Gumble-Morgenstern family (1960)*

If $\theta \in [-1, 1]$, then the function C_θ defined on $[0, 1]^2$ by

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v),$$

is a one-parameter bivariate copula.

Example 4. *Cuadras-Augé family of copulas*

Let $\theta \in [0, 1]$. The function C_θ defined by

$$C_\theta(u, v) = [\min(u, v)]^\theta [uv]^{1-\theta} = \begin{cases} uv^{1-\theta}, & \text{if} u \leq v; \\ u^{1-\theta}v, & \text{if} u \geq v, \end{cases}$$

is another copula function.

Example 5. *The Gaussian Copula*

This copula is simply an application of the defintion of copulas. By taking a multivariate Gaussian distribution function Φ_Σ with mean zero and covariance matrix Σ, and transforming the marginals by the inverse of the standard normal distribution function Φ we obtain the following copula

$$C(x_1, x_2, \ldots, x_d) = \Phi_\Sigma(\Phi^{-1}(x_1), \Phi^{-1}(x_2), \ldots, \Phi^{-1}(x_d)).$$

Example 6. *The t-Copula*

The t-copula is derived in much the same way as the Gaussian copula before. Taking a multivariate centered t-distribution function $t_{\Sigma,\nu}$ with covariance matrix Σ, ν degrees of freedom, and marginal distribution t_ν, we get the following copula

$$C(x_1, x_2, \ldots, x_d) = t_{\Sigma,\nu}(t_\nu^{-1}(x_1), t_\nu^{-1}(x_2), \ldots, t_\nu^{-1}(x_d)).$$

Example 7. *Bivariate Extreme Value copulas*

This family of copulas is obtained from the bivariate extreme-value distribution. Such a copula has the form

$$C_A(u,v) = \exp\left[\log(u,v)A\left\{\frac{\log(u)}{\log(v)}\right\}\right],$$

where A is the dependence function, defined on $[0,1]$. It is assumed to be convex and such that

$$\max(t, 1-t) \leq A(t) \leq 1, \forall t \in [0,1].$$

Example 8. *Archimedean copulas*

A copula is an Archimedean copula if it can be expressed in the form

$$C_\phi(u_1, u_2, \ldots, u_n) = \phi^{-1}\{\phi(u_1) + \phi(u_2) + \ldots + \phi(u_n)\},$$

where $\phi : [0,1] \to [0,\infty)$ is a bijection such that $\phi(1) = 0$, and

$$(-1)^i \frac{d^i}{dx^i}\phi^{-1}(x) > 0, i \in N.$$

ϕ is called the generator of the copula C_ϕ.

As is apparent from the definition, the most important and attractive characteristic of Archimedean copulas is the fact that all information about the n-dimensional dependence structure is contained in the univariate generator ϕ. Thus we can analyze multivariate copulas using a single univariate function.

2.2.5 Generating Copulas

If the examples above do not suffice, or if we want the specific copula for a given pair of random variables and their joint distribution function, we have the option of generating a specific copula for our needs. The concept is flexible enough that we can not only produce copulas for joint distribution functions, but we can also fit them to survival functions and create the so called survival copulas. Below we define joint survival functions and survival copulas.

Definition 2.2.4. For a pair (X, Y) of random variables with joint distribution function H, the joint survival function is defined by

$$\bar{H}(x, y) = P[X > x, Y > y].$$

The marginals of $\bar{H}$ are the functions $\bar{H}(\infty, y)$ and $\bar{H}(x, \infty)$ which are univariate survival functions $\bar{F}$ and $\bar{G}$ where F and G are the distribution functions of X and Y respectively.

Definition 2.2.5. If C is a copula for X and Y, then the survival copula of X and Y is the function $\hat{C} : [0, 1]^2 \to [0, 1]$, given by

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

Also if $\bar{C}$ is the joint survival function for two uniform $(0, 1)$ random variables U and V whose joint distribution function is the copula C, then we have

$$\bar{C}(u, v) = 1 - u - v + C(u, v) = \hat{C}(1 - u, 1 - v).$$

The Inversion Method

Let H be a bivariate distribution function with continuous marginals F and G. A copula C can be constructed by using Sklar's Theorem through the relation,

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)).$$

Using the survival function $\bar{H}$, we can also construct a survival copula by the relation,

$$\hat{C}(u,v) = \bar{H}(\bar{F}^{-1}(u), \bar{G}^{-1}(v)),$$

where $\bar{F}$ and $\bar{G}$ are taken as in Definition 2.2.4

There are several other methods of generating copulas, such as the geometric and algebraic methods, all of which can be found by the interested reader in Nelson's book [18].

2.2.6 Estimating Copulas

Several methods for estimating copulas from a given random sample have been proposed throughout the literature, and below we present a few of the more popular ones, some of which we employed in this thesis. Before proceeding we will define the notation used throughout these algorithms for reference.

Notation and Assumptions

Let the copula we are estimating belong to the family $\{C_\theta, \theta \in \Theta\}$, where Θ is the parameter space. Define the following copula based parametric model for the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$, with cumulative distribution function,

$$F(x;\alpha_1,\alpha_2,\ldots,\alpha_d;\theta) = C(F_1(x_1;\alpha_1), F_2(x_2;\alpha_2), \ldots, F_d(x_d;\alpha_d);\theta),$$

where $F_1, F_2, \ldots, F_d$ are univariate cumulative distribution functions with respective parameters $\alpha_1, \alpha_2, \ldots, \alpha_d$. Let c (mixed derivatives of order d) represent the density of C, and f_j be the marginal density of X_j for $j \in \{1, 2, \ldots, d\}$. Then $\mathbf{X}$ has density function [13]:

$$f(x;\alpha_1,\ldots,\alpha_d;\theta) = c(F_1(x_1;\alpha_1), F_2(x_2;\alpha_2), \ldots, F_d(x_d;\alpha_d);\theta) \prod_{j=1}^{d} f_j(x_j;\alpha_j).$$

Given a sample of size n, $X_1, X_2, \ldots, X_n$ of observations from random vector $\mathbf{X}$, the d log-likelihood functions for the univariate marginals and joint distribution respectively are given by

$$L_j(\alpha_j) = \sum_{i=1}^{n} \log f_j(y_{ij}; \alpha_j), j = 1, 2, \ldots, d,$$

and

$$L(\alpha_1, \alpha_2, \ldots, \alpha_d; \theta) = \sum_{i=1}^{n} \log f(y_i; \alpha_1, \ldots, \alpha_d; \theta),$$

Once one estimates the parameter θ, one has an estimate of the copula.

The Inference Method for Marginals

Presented by Joe and Xu [13], this method consists of performing d separate optimizations of the univariate likelihood functions, followed by an optimisation of the multivariate likelihood as a function of the dependence parameter vector. The steps of the algorithm are summarized below:

1. The log-likelihoods $L_1(\alpha_1), L_2(\alpha_2), \ldots, L_d(\alpha_d)$, of the d univariate marginals are used separately to get estimates $\hat{\alpha_1}, \hat{\alpha_2}, \ldots, \hat{\alpha_d}$ of $\alpha_1, \alpha_2, \ldots, \alpha_d$ respectively;

2. Maximize the function $L(\hat{\alpha_1}, \hat{\alpha_2}, \ldots, \hat{\alpha_d}; \theta)$ over θ to get the estimate $\hat{\theta}$.

Under regularity conditions, $(\hat{\alpha_1}, \hat{\alpha_2}, \ldots, \hat{\alpha_d}, \hat{\theta})$ is the solution to

$$\left(\frac{\partial L_1}{\partial \alpha_1}, \frac{\partial L_2}{\partial \alpha_2}, \ldots, \frac{\partial L_d}{\partial \alpha_d}, \frac{\partial L}{\partial \theta}\right) = 0.$$

This method is useful for models with the closure property of parameters associated with or being expressed in lower-dimensional marginals.

The Maximum Likelihood Method

Presented by Frees and Valdez [8], the method is a natural extension of the maximum likelihood estimate method to copulas and it simply consists of solving the following equation simultaneously

$$\left(\frac{\partial L}{\partial \alpha_1}, \frac{\partial L}{\partial \alpha_2}, \ldots, \frac{\partial L}{\partial \alpha_d}, \frac{\partial L}{\partial \theta}\right) = 0.$$

The Empirical Copula

A natural extension to copulas of the empirical distribution function, this method was rigurously presented by Veraverbeke [31], but a good exposition of it can also be found in Nelson's book [18]. We used this method to estimate copulas in our thesis, as it lends itself naturaly to vector programming found in the R statistical software we employed.

Consider an iid sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, of random vector (X, Y). The bivariate empirical distribution function associated with (X, Y) is

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x, Y_i \leq y\}},$$

with marginals

$$F_n(x) = H_n(x, -\infty) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} \text{ and } G_n(y) = H_n(-\infty, y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \leq y\}},$$

where I_A is the indicator function of the set A. Then the empirical copula function is given by

$$\begin{aligned} C_n(u, v) &= H_n(F_n^{-1}(u), G_n^{-1}(v)) \\ &= \frac{1}{n} \sum_{k=1}^{n} I_{\{X_k \leq F_n^{-1}(u), Y_k \leq G_n^{-1}(v)\}}. \end{aligned}$$

Nelson [18] represented this in a different way, using order statistics, though it is straightforward to see that the two equations are equivalent:

$$C_n(\frac{i}{n}, \frac{j}{n}) = \frac{\text{number of pairs } (x, y) \text{ in the sample with } x \leq x_{(i)}, y \leq y_{(j)}}{n},$$

where $x_{(i)}, y_{(j)}, 1 \leq i, j \leq n$, denote the order statistics of the sample.

Estimating Archimedean Copulas

An archimedean copula is known once one has its generator, a fact that is obvious from its definition. Thus, to estimate an archimedean copula from a given sample of bivariate observations $(X_{11}, X_{21}), (X_{12}, X_{22}), \ldots, (X_{1n}, X_{2n})$, we must identify the generating function ϕ and its afferent parameter α. The algorithm below is restricted to one parameter archimedean copulas, but one can extend the method to multiple parameter archimedean copulas. Genest and Rivest [9] were the first to introduce a straightforward procedure for estimating the generating function ϕ and parameter α. First one considers an intermediate pseudo-observation Z_i (defined in 2.a below), that has distribution function $K(z) = P[Z_i \leq z]$. The authors showed that K is related to an archimedean copula through the relation $K(z) = z - \frac{\phi(z)}{\phi'(z)}$.

1. Estimate Kendall's tau using the usual estimate

$$\tau_n = \binom{n}{2}^{-1} \sum_{i<j} \mathrm{sign}[(X_{1i} - X_{1j})(X_{2i} - X_{2j})].$$

2. (a) Define the pseud-observations Z_i with distribution function $K(z) = P[Z_i \leq z]$ as follows:

$$Z_i = \frac{\{\text{number of } (X_{1j}, X_{2j}) \text{ such that} X_{1j} < X_{1i} \text{ and } X_{2j} < X_{2i}\}}{n-1};$$

   (b) construct the estimate of K_n of K as $K_n(z) = \text{proportion of } Z_i^{'}s \leq z$.

3. Since K has to satisfy the relation

$$K(z) = z - \frac{\phi(z)}{\phi`(z)},$$

we obtain an estimate ϕ_n of ϕ, by solving the equation

$$z - \frac{\phi_n(z)}{\phi_n'(z)} = K_n(z).$$

4. Finally using the estimate for τ given in point 1 above, and the estimate for ϕ obtained in point 3 we can now get an estimate for α by soloving the following equation:

$$\hat{\tau} = 1 + f \int_0^1 \frac{\phi(t)}{\phi'(t)} dt.$$

2.3 Lévy Processes and Copulas

2.3.1 Basic facts about Lévy processes

The following section is a brief summary of classic theory for Lévy processes and infinitely divisible distributions, most of which will be useful for an understanding of Lévy copulas in the following section. The section is made up of a collection of definitions, lemmas and theorems from Cont and Tankov's book [2] on financial modelling, as well as Ross' book [23] on stochastic processes.

Definition 2.3.1. $X = \{X(t)\}_{t \geq 0}$ is said to be a Lévy process if

1. X has independent increments. That is given times a, b, c, d, such that $a < b < c < d$, $X(a) - X(b)$ and $X(c) - X(d)$ are independent.

2. $X(0) = 0$ a.s.

3. X is stochastically continuous (also called continuous in probability or **P**-continuous), if for all $s \geq 0$

$$X(t+s) - X(s) \xrightarrow{P} 0 \text{ as } t \to 0.$$

4. X is time homogeneous, i.e., for $t \geq 0, \mathcal{L}(X(t+s) - X(s))$ does not depend on $s \geq 0$.

5. X is rcll (right continuous with left limits) almost surely. This property is often refered to as càdlàg from the French "continue à droite, limite à gauche".

It is easy to check that both Brownian Motion and Poisson process are Lévy processes.

Proposition 2.3.1. *(Characteristic functions of a Lévy process). Let $(X_t)_{t\geq 0}$ be a Lévy process on R^d. There exists a continuous function $\psi : R^d \to R$ called the characteristic exponent of X, such that*

$$Ee^{iuX(t)} = e^{t\psi(u)}, u \in R^d.$$

Theorem 2.3.1. *The marginal distribution of $X(t)$ is determined by $X(1)$.*

Proof. Since $X(t)$ has independent increments and is time homogenous, we can write

$$\phi_{s+t}(u) = Ee^{iuX(s+t)} = Ee^{iu(X(s+t)-X(t))}e^{iuX(t)}$$

$$= Ee^{iu(X(s+t)-X(t))}Ee^{iuX(t)} = Ee^{iuX(s)}Ee^{iuX(t)} = \phi_s(u)\phi_t(u).$$

The feature of P-continuity implies $\phi_t(u)$ is continuous with respect to t for all u. Proposition 2.3.1 gives us that $\phi_t(u) = e^{t\psi(u)}$, and further $\phi_1(u) = e^{\psi(u)}$. Hence $\phi_t(u) = (\phi_1(u))^t$. □

Definition 2.3.2. Two stochastic processes $\{Z_1(t)\}_{t\in T}$ and $\{Z_2(t)\}_{t\in T}$ are modifications (also called indistinguishable) of each other, if

$$P\{Z_1(t) = Z_2(t)\} = 1 \text{ ,for all } t \in T.$$

Definition 2.3.3. A random vector Y is infinitely divisble (ID) if, for each $n \in N$, there is an i.i.d. sequence $Y_{n,1}, \ldots, Y_{n,n}$ such that

$$Y \stackrel{d}{=} Y_{n,1} + \ldots + Y_{n,n}.$$

Theorem 2.3.2. *(Lévy Processes and Infinite Divisibility). Let $(X_t)_{t \geq 0}$ be a Lévy process. Then, for every t, X_t has an infinitely divisble distribution. Conversely, if F is an infinitely divisible distribution then there exists a Lévy process (X_t) such that the distribution of X_1 is given by F.*

Proof. For a proof of this see Corollary 11.6 in Sato's book [24] on Lévy processes and infinitely divisible distributions. $\square$

Theorem 2.3.3. *(Lévy-Khintchine Formula). Let X be a Lévy process in R^d. There exists a triplet (A, γ, ν), where*

$$\begin{cases} A \text{ a symmetric non-negative definite } d \times d \text{ matrix (the Gaussian covariance)}, \\ \gamma \text{ a constant in } R^d, \\ \nu \text{ a measure on } R^d \text{ with } \nu(0) = 0 \text{ and } \int_{R^d}(|y|^2 \wedge 1)d\nu(y) < \infty \text{ (the Lévy measure).} \end{cases}$$

which in that case is uniquely determined, such that, for all $u \in R^d$ and $t \geq 0$, $Ee^{iuX(t)} = e^{t\psi_u}$, where

$$\psi_u = -\frac{1}{2} < u, Au > +i < u, \gamma > + \int_{R^d} (e^{i<u,y>} - 1 - 1_{(|y| \leq 1}i < u, y >)d\nu(y).$$

Where $< x, y >$ denotes the scalar product of vectors x and y. If $\gamma_0 = \gamma - \int_{|y| \leq 1} y d\nu(y)$ is well-defined and finite, then we may rewrite the Levy-Khintchine Formula with a new triplet $(A, \gamma_0, \nu)_0$ (the drift) as

$$\psi_u = -\frac{1}{2} < u, Au > +i < u, \gamma_0 > + \int_{R^d} (e^{i<u,y>} - 1)d\nu(y).$$

If $\gamma_1 = \gamma + \int_{|y|>1} y d\nu(y)$ *is well-defined and finite, then we can rewrite the Levy-Khintchine Formula with a new triplet* $(A, \gamma_1, \nu)_1$ *(the center) as*

$$\phi_u = -\frac{1}{2} < u, Au > +i < u, \gamma_1 > + \int_{R^d} (e^{i<u,y>} - 1 - i < u, y) d\nu(y).$$

Definition 2.3.4. A compound Poisson process is a Levy process with a generating triplet $(0, 0, \lambda\sigma)_0$, where $\lambda > 0$ is a constant and σ a probability measure on R^d with $\sigma\{0\} = 0$.

Theorem 2.3.4. *Let* $\{N(t)\}_{t\geq 0}$ *be a Poisson process with rate* λ*, and* $\{Y_k\}_{k=1}^{\infty}$ *i.i.d. rv.s, independent of N, with* $\mathcal{L}(Y_k) = \sigma$*, where* $\sigma\{0\} = 0$*. Denoting* $S_0 = 0, S_n = \sum_{k=1}^{n} Y_k$ *for* $n \in N, X(t) = S_{N(t)}$ *is a compound Poisson process with generating triplet* $(0, 0, \lambda\sigma)$*.*

Proof. Rcll sample path and $X(0) \stackrel{d}{=} 0$ are immediate. P-continuity follows from

$$P\{|X(t+s) - X(s)| > \epsilon\} \leq P\{|N(s+t) - N(s)| > 0\} = 1 - e^{-\lambda|t|} \to 0 \text{ as } t \to 0.$$

Independence and homogeneity of increments come by conditioning on the values of N involved. The generating triplet is the claimed one, since

$$Ee^{i<u,X(1)>} = \sum_{n=0}^{\infty} (Ee^{i<u,S_1>})^n \frac{\lambda^n}{n!} e^{-\lambda} = \exp\{\lambda \int_{R^d} (e^{i<u,y>} - 1) d\sigma(y)\}.$$

□

2.3.2 The Lévy Measure and Lévy Copulas

The following section is primarily based on the work of Cont and Tankov published in their book on Financial Models [2]. We reproduce a summary of the principal results and concepts involved and refer the interested reader to their book which gives a comprehensive overview of Lévy copulas.

The law of a Lévy process $\{X_t : t \geq 0\}$ is completely determied by the law of X_t. Thus given fixed time $t = r, r > 0$, the dependence structure of a two-dimensional Lévy process (X_t, Y_t) can be parameterized by the copula C_r, of X_r and Y_r for some $r > 0$. Cont and Tankov [2], however, point out the following issues with this approach:

- **Copulas may be time-dependent**

  "The copula C_r may depend on r. C_s for some $s \neq r$ cannot in general be computed from C_r because C_s also depends on the marginal distributions at time r and at time s." [2]

- **Copulas are invariant for strictly increasing transformations**

  The theorem stating this can be found in Nelsen's book on copulas [18]. However Cont and Tankov point out that the property of infinite divisibility of a random variable is destroyed under strictly increasing transformations. We therefore have that "For given infinitely divisible marginal laws P_{X_t} and P_{Y_t}, it is not clear which copula C_t will yield a two-dimensional infinitely divisible law"[2]. Instead of using copulas, Cont and Tankov suggest modelling the dependence between two Lévy processes in a way that preserves both the dependence structure itself and the dynamic structure of Lévy processes.

Example 9. Dynamic complete dependence for Lévy processes[2]

Let X_t be a pure jump Lévy process. Let Y_t be a Lévy process, constructed from the jumps of $X_t : Y_t = \sum_{s \leq t} \Delta X_s^3$. From the dynamic point of view X_t and Y_t are completely dependent in the following way: The trajectory of one of them can be reconstructed from the other. However, the copula of X_t and Y_t is not that of complete dependence, because Y_t is not a deterministic function of X_t.

Cont and Tankov use this example to argue that in studying the dependence between Lévy processes one should study how their jumps are correlated. In essence processes whose jump times can be inferred from one another are dependent, and this association can be studied relying exclusively on the Lévy measure itself. Below we reproduce two essential results that Cont and Tankov use to buid the Lévy copula theory.

Proposition 2.3.2. *(Abstract margins of Lévy measure). Let* $\boldsymbol{X}_t = (X_t, Y_t)$ *be a two-dimensional Lévy process with generating triplet* (Q, ν, γ)*. Then the component* X_t *of* $\boldsymbol{X}_t$ *has generating triplet* (Q_X, ν_X, γ_X)*, where*

$$Q_X \quad = Q_{11},$$

$$\nu_(B) = \nu(B \times (-\infty, \infty)), \forall B \in \mathcal{B}(R),$$

$$\gamma_X \quad = \gamma_1 + \int_{R^2} x(1_{x^2 \le 1} - 1_{x^2+y^2 \le 1})\nu(dx \times dy).$$

Proposition 2.3.3. *(Independence of Lévy processes). Let* (X_t, Y_t) *be a Lévy process with a Lévy measure* ν *and without a Gaussian coefficient (the positive nonnegative-definite matrix Q). The components of the Lévy process are independent if and only if the support of its measure* ν *is contained in the set* $\{(x, y) : xy = 0\}$*. That is, if and only if the components never jump together. With this restriction*

$$\nu(A) = \nu_X(A) + \nu_Y(A),$$

where $A_X = \{x : (x, 0) \in A\}, A_Y = \{y : (0, y) \in A\}$*, and* ν_X, ν_Y *are Lévy measures of* X_t *and* Y_t *respectively.*

Proof. A proof can be found on page 144 in Cont and Tankov [2] □

2.3.3 Levy copulas for Levy processes with positive jumps

Definition 2.3.5. A d-dimensional abstract *tail integral* is a function $U : [0, \infty]^d \to [0, \infty]$ such that

1. $(-1)^d U$ is a d-increasing function, as defined in point 2 of Definition 2.2.2 above.

2. U is equal to zero if one of its arguments is equal to ∞, U is finite everywhere except at zero, and $U(0,\ldots,0)=\infty$.

The *margins* of a Lévy measure are defined similarly to the margins of a distribution function:

$$U(0,\ldots,x_k,0,\ldots,0)=U_k(x_k).$$

In particular, for every Lévy measure ν on $(0,\infty]\times(0,\infty]$, one can define its tail integral as folows:

$$U(x_1,x_2)=0, \qquad \text{if } x_1=\infty \text{ or } x_2=\infty$$

$$U(x_1,x_2)=\nu([x_1,\infty)\times[x_2,\infty)), \qquad \text{for } (x_1,x_2)\in[0,\infty)\times[0,\infty)\ \{(0,0)\}$$

$$U(0,0) \quad =\infty.$$

Notice that just as we define the tail integral of the Lévy measure, we can just as easily go the other way and define our Lévy measure using its tail integrals because

$$\int_{[0,1]\times[0,1]}|x|^2\nu(dx)=\int_{[0,1]\times[0,1]}|x|^2(dU)<\infty.$$

The exact requirements for a two-dimensional tail integral to define a Lévy measure are specified in the lemma below, reproduced from Chapter 5 in Cont and Tankov [2].

Lemma 2.3.1. *Let U be a two-dimensional tail integral with margins U_1 and U_2. U defines a Lévy measure on $[0,\infty)\times[0,\infty)\setminus\{0,0\}$ if and only if the following condition is met: The margins of U correspond to Lévy measures on $[0,\infty)$. That is, for $k=1,2$*

$$\int_0^1 x^2 dU_k(x)<\infty.$$

Proof. A proof is found on page 147 in Cont and Tankov [2] $\square$

Definition 2.3.6. An n-dimensional positive Lévy copula is a n-increasing, grounded function $F : [0,\infty]^n \to [0,\infty]$, with margins F_k for $k = 1,\ldots,n$, which satisfy $F_k(u) = u$ for all $u \in [0,\infty]$. Recall we explained the concepts of d-increasing and grounded in Definition 2.2.2 above.

The following is a general theorem equivalent to Sklar's theorem for copulas.

Theorem 2.3.5. *Let U be the tail integral of the Lévy measure of an n-dimensional Lévy process with positive jumps and let $U_1,\ldots,U_n$ be the tail integrals of its components. Then there exists an n-dimensional positive Lévy copula F such that, for all vectors $(x_1,\ldots,x_n)$ in R_+^n*

$$U(x_1, x_2, \ldots, x_n) = F(U_1(x_1), \ldots, U_n(x_n)).$$

If the $U_1,\ldots,U_n$ are continuous, then F is unique; otherwise, it is unique on Ran $U_1 \times \ldots \times$Ran U_n. Conversely, if F is an n-dimensional positive Lévy copula and $U_1,\ldots,U_n$ are tail integrals on $(0,\infty)$, then the function U defined above is the tail integral of an n-dimensional Lévy process with positive jumps having marginal tail integrals $U_1,\ldots,U_n$.

Proof. A proof of the two-dimensional case can be found on page 148 in Cont and Tankov [2] with general guidelines for the n-dimensional case given on page 155. $\square$

Proposition 2.3.4. *Let C be a 2-copula (not a Lévy copula). Let $f(x)$ be an increasing convex function from $[0,1]$ to $[0,\infty]$. Then*

$$F(x,y) = f(C(f^{-1}(x), f^{-1}(y)))$$

defines a two-dimensional positive Lévy copula.

Proof. Given on page 153 in Cont and Tankov [2]. □

Proposition 2.3.5. *Let ϕ be a strictly decreasing function from $[0,\infty]$ to $[0,\infty]$ such that $\phi(0)=\infty$ and $\phi(\infty)=0$. Let the quasi-inverse $\phi^{(-1)}$ have derivatives up to the order n on $(0,\infty)$ with alternating signs. That is $(-1)^k \frac{d^k\phi^{-1}(r)}{dr^k} > 0$. Then*

$$F(x_1,\ldots,x_n) = \phi^{(-1)}(\phi(x_1)+\ldots+\phi(x_n)),$$

defines an n-dimensional positive Levy copula.

Proof. A proof of the two-dimensional case is given on page 153 in Cont and Tankov [2] with general guidlines for the n-dimensional case given on page 155. □

For instance if we take $\phi(u) = u^{-\theta}$ for $\theta > 0$ we get a family of positive Levy copulas of the form

$$F_\theta(u,v) = (u^{-\theta}+v^{-\theta})^{-1/\theta}.$$

This is the family of positive Clayton Lévy copulas, since they resemble the Clayton copulas obtained by taking $\phi(r) = (r^{-\theta}-1)^{-\frac{1}{\theta}}$ as the generator of the Archimedean copulas defined in Example 8 Section 2.2.4. Notice that for this Lévy copula the limiting cases $\theta\to\infty$ and $\theta\to 0$ correspond to complete dependence and independence, respectively. Positive Lévy couplas also have a probabilistic interpretation, as stated in the following results from Cont and Tankov [2].

Lemma 2.3.2. *Let F be a two-dimensional positive Lévy copula. Then for almost all $x\in[0,\infty)$, the function*

$$F_x(y) = \frac{\partial}{\partial x}F(x,y)$$

exists and is continuous for all $y\in[0,\infty]$. Moreover, it is a distribution function of a positive random variable, that is, it is increasing and satisfies $F_x(0)=0$ and $F_x(\infty)=1$.

Proof. A proof of the above result is given on page 154 of Cont and Tankov [2]. □

Theorem 2.3.6. *Let (X_t, Y_t) be a two-dimensional Lévy process with jumps in the positive quadrant, having marginal tail integrals U_1, U_2 and Lévy copula F. Let ΔX_t and ΔY_t be the jump sizes of the two components at time t. Then, if U_1 has non-zero density at x, $F_{U_1(x)}$ is the distribution function of $U_2(\Delta Y_t)$ conditionally on $\Delta X_t = x$:*

$$F_{U_1(x)}(y) = P\{U_2(\Delta Y_t) \leq y | \Delta X_t = x\}.$$

Proof. A reference to a proof is given on page 155 in Cont and Tankov [2]. □

2.3.4 Lévy 2-copulas for general Lévy processes

Cont and Tankov use the following notation $\bar{\mathbf{R}} := \mathbf{R} \cup \{-\infty, \infty\}$ and $\mathbf{R}_+$ as the set of non-negative real numbers and we use the same notation throughout this section.

Definition 2.3.7. A function $F : \bar{\mathbf{R}}^2 \to \bar{\mathbf{R}}$ is called a general Lévy 2-copula if

- F is 2-increasing,
- $F(0, x) = F(x, 0) = 0$ for all x,
- $F(x, \infty) - F(x, -\infty) = F(\infty, x) - F(-\infty, x) = x$.

An example of a general Lévy 2-copula is given below.

Example 10. (General Clayton Lévy 2-copula)

$$F_\theta(u, v) = \begin{cases} (|u|^{-\theta} + |v|^{-\theta})^{-1/\theta} 1_{xy \geq 0}, & \text{for } \theta > 0; \\ -(|u|^{-\theta} + |v|^{-\theta})^{-1/\theta} 1_{xy \leq 0}, & \text{for } \theta < 0. \end{cases}$$

The tail integral of a general Lévy measure on $\mathbf{R}$ is defined as follows:

Definition 2.3.8. Let ν be a Lévy measure on $\mathbf{R}$. The *tail integral* of ν is a function $U : \bar{\mathbf{R}} \setminus 0 \to [-\infty, \infty]$ defined by:

$$U(x) = \nu([x, \infty]), \text{ for } x \in (0, \infty),$$

$$U(x) = -\nu((-\infty, -x]), \text{ for } x \in (-\infty, 0),$$

$$U(\infty) = U(-\infty) = 0.$$

Using Lévy 2-copulas for general Lévy processes, we can use sufficiently smooth general Lévy copulas to construct two-dimensional Lévy densities from one-dimensional ones

Proposition 2.3.6. *Let F be a two-dimensional Lévy copula, continuous on $[-\infty, \infty]^2$, such that $\frac{\partial^2 F(u,v)}{\partial u \partial v}$ exists on $(-\infty, \infty)^2$ and let U_1 and U_2 be one-dimensional tail integrals with Lévy measures ν_1 and ν_2. Then*

$$\left.\frac{\partial^2 F(u,v)}{\partial u \partial v}\right|_{u=U_1(x), v=U_2(y)} \nu_1(x)\nu_2(y)$$

is the Lévy density of a Lévy measure, with marginal Levy measures ν_1 and ν_2.

Proof. This result is stated on page 157 in Cont and Tankov [2]. □

To construct Lévy copulas with both positive and negative jumps we simply treat each corner of the Lévy measure separately, as in the definition below.

A method for constructing general tail integrals: Consider the 1-dimensional case. Let ν be a Lévy measure on $\mathbf{R}$. This measure has two tail integrals, $U^+ : [0, \infty] \to [0, \infty]$ for the positive part and $U^- : [-\infty, 0] \to [-\infty, 0]$ for the negative part, defined as follows:

$$U^+(x) = \nu([x, \infty)), \qquad \text{for } x \in (0, \infty), U^+(0) = \infty, U^+(\infty) = 0;$$

$$U^-(x) = \nu((-\infty, x]), \text{ for } x \in (-\infty, 0), U^-(0) = -\infty, U^+(-\infty) = 0.$$

Now, consider the 2-dimensional case. Let ν be a Lévy measure on $\mathbf{R}^2$ with marginal tail integrals $U_1^+, U_1^-, U_2^+, U_2^-$. This measure has four tail integrals: U^{++}, U^{+-}, U^{-+}, and U^{--}, where each tail integral is defined on its respective quadrant, including the coordinate axis, as follows:

$$\begin{aligned}
U^{++}(x, y) &= \nu([x, \infty) \times [y, \infty)), && \text{if } x \in (0, \infty) \text{ and } y \in (0, \infty); \\
U^{+-}(x, y) &= -\nu([x, \infty) \times (-\infty, y]), && \text{if } x \in (0, \infty) \text{ and } y \in (-\infty, 0); \\
U^{-+}(x, y) &= -\nu((-\infty, x] \times [y, \infty)), && \text{if } x \in (-\infty, 0) \text{ and } y \in (0, \infty); \\
U^{--}(x, y) &= \nu((-\infty, x] \times (-\infty, y]), && \text{if } x \in (-\infty, 0) \text{ and } y \in (-\infty, 0).
\end{aligned}$$

If x or y is equal to $+\infty$, or $-\infty$, the corresponding tail integral is zero. If x or y is equal to zero, the tail integrals satisfy the following "margin" conditions:

$$U^{++}(x, 0) - U^{+-}(x, 0) = U_1^+(x),$$

$$U^{-+}(x, 0) - U^{--}(x, 0) = U_1^-(x),$$

$$U^{++}(0, y) - U^{-+}(0, y) = U_2^+(y),$$

$$U^{+-}(0, y) - U^{--}(0, y) = U_2^-(y).$$

Taking the two-dimensional Lévy measure, separately for each quadrant, we may state the equivalent of Sklar's Theorem for general Lévy measures as follows:

Theorem 2.3.7. *Let ν be a Lévy measure on $\mathbf{R}^2$ with marginal tail integrals U_1^+, U_1^-, U_2^+, and U_2^-. Then there exists a Lévy copula F such that U^{++}, U^{+-}, U^{-+} and U^{--} are* *tail integrals of ν, as follows:*

$$U^{++}(x,y) = F(U_1^+(x), U_2^+(y)), \quad \text{if } x \geq 0 \text{ and } y \geq 0;$$
$$U^{+-}(x,y) = F(U_1^+(x), U_2^-(y)), \quad \text{if } x \geq 0 \text{ and } y \leq 0;$$
$$U^{-+}(x,y) = F(U_1^-(x), U_2^+(y)), \quad \text{if } x \leq 0 \text{ and } y \geq 0;$$
$$U^{--}(x,y) = F(U_1^-(x), U_2^-(y)), \quad \text{if } x \leq 0 \text{ and } y \leq 0.$$

If the marginal tail integrals are absolutely continuous[1] *and ν does not change the coordinate axes, the Lévy copula is unique. Conversely, if F is a Lévy copula and $U_1^+, U_1^-, U_2^+, U_2^-$ are tail integrals of one-dimensional Lévy measures, then the above formulas define a set of tail integrals of a two-dimensional Lévy measure.*

Proof. A proof is given on page 158 in Cont and Tankov [2]. □

As a technique to construct general Lévy copulas, Cont and Tankov suggest getting them "from positive ones by gluing them together", which "amounts to specifying the dependence of different signs separately" [2].

By this procedure, letting $F^{++}, F^{--}, F^{-+}, F^{+-}$ be positive Lévy copulas, it can be shown that ([2] p.160)

$$F(x,y) = F^{++}(c_1|x|, c_2|y|)1_{x\geq 0, y\geq 0} + F^{--}(c_3|x|, c_4|y|)1_{x\leq 0, y\leq 0}$$

$$-F^{+-}((1-c_1)|x|, (1-c_4)|y|)1_{x\geq 0, y\leq 0} - F^{-+}((1-c_3)|x|, (1-c_2)|y|)1_{x\leq 0, y\geq 0}$$

defines a Lévy copula if $c_1, \ldots, c_4$ are constants between 0 and 1. Such Lévy copulas are refered to as *constant proportion Lévy copulas.* One can similarly define general n-dimensional Lévy copulas, as follows:

Definition 2.3.9. An n-dimensional Lévy copula is a function $F : \bar{\mathbf{R}}^n \to \bar{\mathbf{R}}$ with the

[1] A function $F : \mathbf{R} \to \mathbf{R}$ is called *absolutely continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that for any finite set of disjoint intervals $(a_1, b_1), (a_2, b_2) \ldots (a_N, b_N), \sum_{j=1}^{N}(b_j - a_j) < \delta$ implies that $\sum_{j=1}^{N} |F(b_j) - F(a_j)| < \epsilon$.

following three properties:

- F is n-increasing,
- F is equal to zero if at least one of its arguments is zero and,
- $F(x, \infty, \ldots, \infty - F(x, -\infty, \ldots, -\infty) = x$

  $F(\infty, x, \infty, \ldots, \infty) - F(-\infty, x, -\infty, \ldots, -\infty) = x,$ etc.

Using a special type of interval

$$\mathcal{I}(x) = \begin{cases} [x, \infty), & \text{if } x > 0; \\ (-\infty, x], & \text{if } x < 0. \end{cases}$$

tail integrals $U_1, \ldots, U_n$ of the Lévy measure can be computed everywhere except on the axes as follows:

$$\nu(\mathcal{I}(x_1)\times, \ldots, \times\mathcal{I}(x_n)) =$$

$$(-1)^{\text{sign}(x_1)\ldots\text{sign}(x_n)} F(U_1^{\text{sign}(x_1)}(x_1), \ldots, U_n^{\text{sign}(x_n)}(x_n)).$$

Here $x_k \in \mathbf{R} \setminus \{0\}$.

Constructing Lévy copulas in higher dimensions using the constant proportion methods not very practical. In the general case of 2^n positive Lévy copulas, a large number of constants must be specified. In such situations, it is better to use a simplified construction such as an n-dimensional Clayton Lévy copula.

3 Measures of Dependence for Random Variables

3.1 Szekely's Correlation

This measure of dependence was introduced by Szekely, Rizzo and Bakirov in their 2007 paper [30]. The measure was given the uninspiring name of *distance correlation*, which can be somewhat misleading. In essence, Szekely et. al. try to measure the correlation between two random vectors by taking the L^2 norm with a singular weight function of the distance between joint characteristic functions and the product of marginal characteristic functions. As we have done thus far in the thesis, we will stick to the original paper's notations and begin by introducing these, before moving to the definitions and theorems related to this measure of dependence.

Throughout this section, X and Y are random vectors, with values in $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively, where p and q are positive integers. Their joint and marginal characteristic functions are denoted, respectively, as $f_{X,Y}$, f_X, and f_Y. The scalar product of vectors t and s is denoted by $< t, s >$, and for complex valued functions $f()$, the complex conjugate of f is denoted by $\bar{f}$, and $|f|^2 = f\bar{f}$. The Euclidean norm of x in $\mathbf{R}^p$ is $|x|_p$. The paper then defines the measure of dependence of interest as follows.

Definition 3.1.1. (*Distance correlation*). The distance correlation (dCor) between random vectors X and Y in $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively, with finite first moments is the nonnegative number $R(X, Y)$ defined by

$$R(X,Y) = \begin{cases} \frac{V^2(X,Y)}{\sqrt{V^2(X)V^2(Y)}}, & V^2(X)V^2(Y) > 0; \\ 0, & V^2(X)V^2(Y) = 0. \end{cases}$$

Here $V^2(X, Y)$ is the so called distance covariance, suggesting an analogy to the classic covariance concept, and is defined as follows:

$$V^2(X,Y) = \frac{1}{c_p c_q} \int_{R^{p+q}} \frac{|f_{X,Y}(t,s) - f_X(t) f_Y(s)|^2}{|t|_p^{1+p} |s|_q^{1+q}} dtds,$$

with

$$c_d = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}, \forall d \in Z, d \geq 1.$$

Similarly Szekely et al. define the distance Variance as follows

$$V^2(X) = \frac{1}{c_p^2} \int_{R^{2p}} \frac{|f_{X,X}(t,s) - f_X(t) f_X(s)|^2}{|t|_p^{1+p} |s|_p^{1+p}} dtds,$$

where

$$f_{X,X}(t,s) = f_X(t+s).$$

The format of the distance correlation definition is chosen on purpose by the authors to suggest an analogy with the classical correlation coefficient. Indeed they show that this new measure of dependence does indeed have similar properties with Pearson's product moment correlation, which they prove in the theorem we reproduce below.

Theorem 3.1.1. *(Properties of dCor).*

1. *If* $E(|X|_p + |Y|_q) < \infty$*, then* $0 \leq R \leq 1$*, and* $R(X,Y) = 0$ *if and only if* X *and* Y *are independent.*

2. $0 \leq R \leq 1$.

3. *If* $R(X,Y) = 1$*, then there exists a vector* a*, a nonzero real number* b *and an orthogonal matrix* C *such that* $Y = a + bXC$*.*

Proof. The proof can be found in Szekely et al. paper [30]. □

3.1.1 Extending to the Multivariate Case and Loosening the Moment Conditions

As we have already seen at its core, this new measure of dependence is simply the weighted L^2 norm of the difference between the joint characteristic function of two random variables and their individual characteristic funtions. A natural extension of this to d dimensions is to then take the absolute value of the difference between the d-dimensional joint characteristic function and the product of the individual marginals weighted by a similar weight function as chosen by Szekely et al. So we propose the following distance correlation between n random vectors $X_1, \dots, X_n$, taking values in $\mathbf{R}^{p_1}, \dots, \mathbf{R}^{p_2}$, respectively, and having finite moments of order α $(0 < \alpha < 2)$:

$$R(X_1, \dots, X_d) = \begin{cases} \frac{V^2(X_1,\dots,X_d)}{\sqrt{V^2(X_1)\dots V^2(X_d)}}, & V^2(X_1)\dots V^2(X_d) > 0; \\ 0, & V^2(X_1)\dots V^2(X_d) = 0, \end{cases}$$

where

$$V^2(X_1, \dots, X_d) = \|f_{X_1,\dots,X_d}(t_1, \dots, t_d) - f_{X_1}(t_1)\dots f_{X_d}(t_d)\|^2$$

$$= \frac{1}{c_{p_1}\dots c_{p_d}} \int_{R^{p_1+\dots+p_d}} \frac{|f_{X_1\dots X_d}(t_1, \dots, t_d) - f_{X_1}(t_1)\dots f_{X_d}(t_d)|^2}{|t_1|_{p_1}^{\alpha+p_1}\dots|t_d|_{p_d}^{\alpha+p_d}} dt_1\dots dt_d.$$

Before setting out to show this new definition has the same properties as the one introduced in Szekely et al. for two dimensions, we will make use of a key result from Szekely's paper[30].

Lemma 3.1.1. If $0 < \alpha < 2$, *then for all* x *in* $\mathbf{R}^d$

$$\int_{R^d} \frac{1 - cos\langle t, x\rangle}{|t|_d^{d+\alpha}} dt = C(d, \alpha)|x|^\alpha,$$

where

$$C(d,\alpha) = \frac{2\pi^{d/2}\Gamma(1-\alpha/2)}{\alpha 2^{\alpha}\Gamma((d+\alpha)/2)}.$$

Proof. Introduce

$$A = \int_{R^{d-1}} \frac{dz_2 dz_3 \ldots dz_d}{(1+z_2^2+z_3^2+\ldots+z_d^2)^{\frac{d+\alpha}{2}}}.$$

Then by the formulas 3.3.2.1, p.585, 2.2.4.24 p.298 and 2.5.3.13 p.387 of Prudnikov, Brychkov and Marichev (1986) [21], we have

$$A = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^{\infty} \frac{x^{d-2}dx}{(1+x^2)^{\frac{d+\alpha}{2}}} = \frac{\pi^{\frac{d-1}{2}}\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{d+\alpha}{2}\right)},$$

and

$$\frac{d}{da}\left(\int_0^{\infty} \frac{1-\cos ax}{x^{1+\alpha}}dx\right) = a^{\alpha-1}\int_0^{\infty} \frac{\sin x}{x^{\alpha}}dx$$

$$= a^{\alpha-1}\frac{\sqrt{\pi}\Gamma\left(1-\frac{\alpha}{2}\right)}{2^{\alpha}\Gamma\left(\frac{\alpha+1}{2}\right)}.$$

Introduce the new variables $s_1 := z_1, s_k := s_1 z_k$ for $k \geq 2$. Then

$$C(d,\alpha) = A \times \int_{-\infty}^{\infty} \frac{1-\cos z_1}{|z_1|^{1+\alpha}}dz_1$$

$$= \frac{\pi^{\frac{d-1}{2}}\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{d+\alpha}{2}\right)} \times \frac{2\sqrt{\pi}\Gamma\left(1-\frac{\alpha}{2}\right)}{\alpha 2^{\alpha}\Gamma\left(\frac{\alpha+1}{2}\right)}$$

$$= \frac{2\pi^{\frac{d}{2}}\Gamma\left(1-\frac{\alpha}{2}\right)}{\alpha 2^{\alpha}\Gamma\left(\frac{d+\alpha}{2}\right)} > 0.$$

□

For finiteness of $\|f_{X_1,\ldots,X_d}(t_1,\ldots,t_d) - f_{X_1}(t_1)\ldots f_{X_d}(t_d)\|^2$, the weighted L^2 norm mentioned above, it is sufficient that $E|X_i|_p^{\alpha} < \infty, \forall i \in \{0,1,\ldots,d\}$. By the Cauchy-

Bunyakovsky inequality,

$$|f_{X_1\dots X_d}(t_1,\dots,t_d)-f_{X_1}(t_1)\dots f_{X_d}(t_d)|^2 =$$

$$= |E\left(e^{i<t_1,X_1>}\dots e^{i<t_d,X_d>} - f_{X_1}(t_1)\dots f_{X_d}(t_d)\right)|^2$$

$$= |E\Big(e^{i<t_1,X_1>}\dots e^{i<t_d,X_d>} - 2^{d-1}f_{X_1}(t_1)\dots f_{X_d}(t_d)$$

$$+(2^{d-1}-1)f_{X_1}(t_1)\dots f_{X_d}(t_d)\Big)|^2$$

$$= |E\left(e^{i\langle t_1,X_1\rangle} - f_{X_1}(t_1)\right)\dots\left(e^{i\langle t_d,X_d\rangle} - f_{X_d}(t_d)\right)|^2$$

$$\leq E|e^{i\langle t_1,X_1\rangle} - f_{X_1}(t_1)|^2\dots E|e^{i\langle t_d,X_d\rangle} - f_{X_d}(t_d)|^2$$

$$= E|1-e^{-itX_1}f_{X_1}(t_1)|^2\dots E|1-e^{-itX_d}f_{X_d}(t_d)|^2$$

$$= E\left(1-e^{-itX_1}Ee^{itX_1}\right)\left(1-e^{itX_1}Ee^{-itX_1}\right)$$

$$\dots E\left(1-e^{-itX_d}Ee^{itX_d}\right)\left(1-e^{itX_d}Ee^{-itX_d}\right)$$

$$= \left(1-2|f_{X_1}(t_1)|^2+|f_{X_1}(t_1)|^2\right)\dots\left(1-2|f_{X_2}(t_2)|^2+|f_{X_d}(t_d)|^2\right)$$

$$= \left(1-|f_{X_1}(t_1)|^2\right)\dots\left(1-|f_{X_d}(t_d)|^2\right).$$

Furthermore if $E\left(|X_1|^{\alpha}_{p_1}+\dots+|X_d|^{\alpha}_{p_d}\right) < \infty$, then by Lemma 1 and Fubini's theorem it follows that

$$\frac{1}{c_{p_1}\dots c_{p_d}}\int_{R^{p_1+\dots+p_d}} \frac{|f_{X_1\dots X_d}(t_1,\dots,t_d)-f_{X_1}(t_1)\dots f_{X_d}(t_d)|^2}{|t_1|^{\alpha+p_1}_{p_1}\dots|t_d|^{\alpha+p_d}_{p_d}}dt_1\dots dt_d$$

$$\leq \int_{R^{p_1}}\frac{1-|f_{X_1}(t_1)|^2}{c_{p_1}|t|^{\alpha+p_1}_{p_1}}dt_1\dots\int_{R^{p_d}}\frac{1-|f_{X_d}(t_d)|^2}{c_{p_d}|t_d|^{\alpha+p_d}_{p_d}}dt_d =$$

$$= E\left[\int_{R^{p_1}}\frac{1-cos\langle t,X_1-X_1'\rangle}{c_{p_1}|t_1|^{\alpha+p_1}_{p_1}}dt_1\right]\dots E\left[\int_{R^{p_d}}\frac{1-cos\langle t_d,X_d-X_d'\rangle}{c_{p_d}|t_d|^{\alpha+p_d}_{p_d}}dt_d\right] =$$

$$= E|X_1-X_1'|^{\alpha}_{p_1}\dots E|X_d-X_d'|^{\alpha}_{p_d} \leq \infty.$$

Finally we are ready to show the main result of interest.

Theorem 3.1.2. If $E(|X_1|^{\alpha}_{p_1} + \ldots + |X_d|^{\alpha}_{p_d}) < \infty$, then $0 \leq R \leq 1$, and

$R(X_1, \ldots, X_d) = 0$ if and only if $\mathrm{X}_1 \ldots X_d$ are independent.

Proof. $R(X_1, \ldots, X_d)$ exists whenever $X_1, \ldots, X_d$ have finite moments of order α and $X_1, \ldots, X_d$ are independent if and only if the numerator

$$V^2(X_1, \ldots, X_d) = \|f_{X_1 \ldots X_d}(t_1, \ldots, t_d) - f_{X_1}(t_1) \ldots f_{X_d}(t_d)\|^2,$$

of $R^2(X_1, \ldots, X_d)$ is zero. Let $U_j = e^{i\langle t_j, X_j \rangle} - f_{X_j}(t_j), j \in \{1, 2, \ldots, d\}$. Then

$$|f_{X_1,\ldots,X_d}(t_1, \ldots, t_d) - f_{X_1} \ldots f_{X_d}(t_d)|^2 =$$

$$= |E[U_1 \ldots U_d]|^2 \leq (E[|U_1| \ldots |U_d|])^2 \leq$$

$$\leq E[|U_1|^2 \ldots |U_d|^2] = \left(1 - |f_{X_1}(t_1)|^2\right) \ldots \left(1 - |f_{X_d}(t_d)|^2\right).$$

□

3.1.2 Empirical Calculations for Distance Correlation

Extending calculations for the empirical distance correlation to d-dimensions is a straightforward matter of applying the same technique used by Szekely et. al. in their paper. We simply take the joint and marginal empirical characteristic functions defined as follows:

$$f^n_{X_1,\ldots,X_d}(t_1, \ldots, t) = \frac{1}{n} \sum_{k=1}^{n} \exp\{i\left(\langle t_1, X_{1k} \rangle + \ldots + \langle t_d, X_{dk} \rangle\right)\},$$

$$f^n_{X_j}(t_j) = \frac{1}{n} \sum_{k=1}^{n} \exp\{i\langle t_j, X_{jk} \rangle\}, \forall j \in \{1, 2, \ldots, d\}.$$

Using this we have the following result for empirical distance correlation.

Definition 3.1.2. If $(X_1, \ldots, X_d)$ is a sample from d-dimensional random vector $\mathbf{X}$, with joint characteristic function $f_{X_1 \ldots X_d}(t_1, \ldots, t_d)$ and marginal characteristic functions $f_{X_i}(t_i)$, where $i \in 1, \ldots, d$, then the empirical distance covariance and empirical distance variance are defined as follows:

$$
\begin{aligned}
V_n^2(X_1, \ldots, X_d) &= \| f^n_{X_1 \ldots X_d}(t_1, \ldots, t_d) - f^n_{X_1}(t_1) \ldots f^n_{X_d}(t_d) \|^2 \\
&= \int_{R^{p_1 + \ldots + p_d}} \frac{| f^n_{X_1 \ldots X_d}(t_1, \ldots, t_d) - f^n_{X_1}(t_1) \ldots f^n_{X_d}(t_d) |^2}{c_{p_1} \ldots c_{p_d} |t_1|_{p_1}^{\alpha + p_1} \ldots |t_d|_{p_d}^{\alpha + p_d}} dt_1 \ldots dt_d, \\
V_n^2(X_i) &= \| f^n_{X_i}(t_1 + \ldots + t_d) - f^n_{X_i}(t_1) \ldots f^n_{X_i}(t_d) \|^2 \\
&= \int_{R^{p_1 + \ldots + p_d}} \frac{| f^n_{X_i}(t_1 + \ldots + t_d) - f^n_{X_i}(t_1) \ldots f^n_{X_i}(t_d) |^2}{c_{p_1} \ldots c_{p_d} |t_1|_{p_1}^{\alpha + p_1} \ldots |t_d|_{p_d}^{\alpha + p_d}} dt_1 \ldots dt_d.
\end{aligned}
$$

Recall that $c_{p_i} = \frac{2\pi^{p_i/2}\Gamma(1-\alpha/2)}{\alpha 2^\alpha \Gamma((p_i+\alpha)/2)}$ is a function of α and p_i.

Theorem 3.1.3. If $E|X_j|_{p_j}^\alpha < \infty, \forall j \in \{1, 2, \ldots, d\},$ then almost surely

$$\lim_{n \to \infty} V_n(X_1, \ldots, X_d) = V(X_1, \ldots, X_d).$$

Proof. The fact that

$$\lim_{n \to \infty} | f^n_{X_1 \ldots X_d}(t_1, \ldots, t_d) - f^n_{X_1}(t_1) \ldots f^n_{X_d}(t_d) |^2 =$$

$$= | f_{X_1 \ldots X_d}(t_1, \ldots, t_d) - f_{X_1}(t_1) \ldots f_{X_d}(t_d) |^2$$

follows trivially from the definition of empirical characteristic functions, by way of the following equality

$$\lim_{n \to \infty} f^n_{X_1 \ldots X_d}(t_1, \ldots, t_d) = f_{X_1 \ldots X_d}(t_1, \ldots, t_d).$$

Notice that $f^n_{X_1, \ldots, X_d}(t_1, \ldots, t_d)$ and $f^n_{X_j}(t_j)$ are bounded for all j, so by the Lebesgue dominated convergence theorem we can conclude the following:

$$\lim_{n\to\infty} \|f^n_{X_1\dots X_d}(t_1,\dots,t_d) - f^n_{X_1}(t_1)\dots f^n_{X_d}(t_d)\|^2 =$$

$$= \lim_{n\to\infty} \int_{R^{p_1+\dots+p_d}} \frac{|f^n_{X_1\dots X_d}(t_1,\dots,t_d) - f^n_{X_1}(t_1)\dots f^n_{X_d}(t_d)|^2}{c_{p_1}\dots c_{p_d}|t_1|^{1+p_1}_{p_1}\dots|t_d|^{1+p_d}_{p_d}} dt_1\dots dt_d$$

$$= \int_{R^{p_1+\dots+p_d}} \frac{\lim_{n\to\infty} |f^n_{X_1\dots X_d}(t_1,\dots,t_d) - f^n_{X_1}(t_1)\dots f^n_{X_d}(t_d)|^2}{c_{p_1}\dots c_{p_d}|t_1|^{1+p_1}_{p_1}\dots|t_d|^{1+p_d}_{p_d}} dt_1\dots dt_d$$

$$= \int_{R^{p_1+\dots+p_d}} \frac{|f_{X_1\dots X_d}(t_1,\dots,t_d) - f_{X_1}(t_1)\dots f_{X_d}(t_d)|^2}{c_{p_1}\dots c_{p_d}|t_1|^{1+p_1}_{p_1}\dots|t_d|^{1+p_d}_{p_d}} dt_1\dots dt_d$$

$$= \|f_{X_1\dots X_d}(t_1,\dots,t_d) - f_{X_1}(t_1)\dots f_{X_d}(t_d)\|^2.$$

This completes our proof. □

3.2 Schweizer and Wolff Sigma

Scheweizer and Wolff introduced this dependency measure in their 1981 paper [25]. The measure was an elegant modification of Spearman's rank rho using copulas and we present it below as one of the two principal measures of dependence for random variables studied in this thesis:

$$\sigma_{X,Y} = 12 \int_I \int_I |C(u,v) - uv| dudv.$$

Nelson [18] points out that any suitably normalized measure of distance between the surfaces $C(u,v)$ and uv, i.e. any L_p distance, should yield a symmetric nonparametric measure of dependence. For any p, $1 \le p < \infty$, the L_p distance between general copula C and the independence copula $\Pi = uv$ is given by

$$\left(k_p \int_I \int_I |C(u,v) - uv|^p dudv\right)^{1/p},$$

where k_p is a constant such that this quantity is equal to 1 when C is the minimum or maximum copulas. For our purpose we used the original measure developed by Schweizer and Wolff both for simplicity and because a higher L_p distance would not provide us with any discernable advantages for studying dependence between random variables.

3.2.1 Extension to the Multivariate Case

Extending Schweizer and Wolff's Sigma to d-dimensions is relatively straightforward. Sklar's theorem is easily extended to d-dimensions [28], and the results hold as well as in the bivariate case. Furthemore we still have the same tools for generating copulas as in the bivariate case, thus we define a d-dimensional multivariate copula as follows

$$C(u_1, \ldots, u_d) = H\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right),$$

where $H(x_1, \ldots, x_d)$ is the multivariate joint distribution of the d random vectors we are modeling, and $F_1, \ldots, F_d$ are their respective univariate marginal distributions.

In addition we have the exact same independence copula as in the bivariate case, with the appropriate modifications as follows[18]:

$$\Pi^d = u_1 \cdot \ldots \cdot u_d.$$

With these definitions in place we can then easily extend our measure of dependence to d-dimensions, following the work done by Wolf in 1981 [33], by

$$\sigma_C^d = \frac{2^d(d+1)}{2^d - (d+1)} \int \ldots \int_{I^d} |C(u_1, \ldots, u_d) - u_1 \times \ldots \times u_d| du_1 \ldots du_d,$$

where the factor $\frac{2^d(d+1)}{2^d-(d+1)}$ is order to make the dependence measure take values betwen 0 and 1. Indeed,

$$\int \cdots \int_{I^d} |C(u_1, \ldots, u_d) - u_1 \times \ldots \times u_d| du_1 \ldots du_d \leq \frac{2^d - (d+1)}{2^d(d+1)},$$

which can be trivially be proven by replacing $C(u_1, \ldots, u_d)$ with the maximum multivariate copula $M(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d)$ and computing the integral above. See (Wolff [33]) for details.

3.2.2 Empirical copulas and dependence

The dependence measures discussed above are designed to work with given marginal and joint distribution functions. However, one may be dealing with a random sample from an unknown joint distribution function, with an as of yet an undetermined copula function. To this end, Nelson [18] proposed the concept of an emprical copula in his book. We will reproduce this concept below, then we will extend it to the multidimensional case and use it to calculate the empirical version of Schweizer and Wolff's sigma.

Definition 3.2.1. (*Empirical Copula*:) Let $\{(x_k, y_k\}_{k=1}^n$ denote a sample of size n from a continuous bivariate distribution. The *empirical copula* is the function C_n given by

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{number of pairs } (x, y) \text{ in the sample with } x \leq x_{(i)}, y \leq y_{(j)}}{n},$$

where $x_{(i)}$ and $y_{(j)}, 1 \leq i, j \leq n$, denote order statistics from the sample.

This leads us to the natural empirical extension for Schweizer and Wolff's sigma reproduced once more from Nelson's book [18] as

$$\sigma = \frac{12}{n^2 - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| C_n\left(\frac{i}{n}, \frac{j}{n}\right) - \frac{i}{n}\frac{j}{n} \right|.$$

Extending this to the multidimensional case requires a few more definitions and the use of empirical distribution functions as follows.

Definition 3.2.2. (*Inverse Distribution Function:*) Let $\mathbf{X}_i = (X_{1i}, \ldots, X_{di}), i = 1, 2, \ldots,$ be an independent sample from a d-dimensional random vector $\mathbf{X} \in \mathbf{R}^d$ with distribution function

$$\mathbf{H}(x_1, \ldots, x_d) := P(X_1 \leq x_1, \ldots, X_d \leq x_d),$$

and marginal df's $F_i = P(X_i \leq x_i)$. Then the inverse of the marginal df's F_i for $i = 1, \ldots, d$ is given by

$$F_i^{-1}(u_i) = \inf\{x_i : F_i(x_i) \geq u_i\},$$

for $0 < u_i < 1$ and $i = 1, \ldots, d$. Similarly, we define the copula function of $\mathbf{H}$ as

$$\mathbf{C}(u_1, \ldots, u_d) = P(F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d),$$

for every $\mathbf{u} \in [0, 1]^d$.

By Sklar's Theorem[28], we know that this copula exists, and it is uniquely determined if the F_i's are continous, and uniquely determined on Ran $F_1 \times \ldots \times$ Ran F_d. Furthermore, the copula is given by

$$\mathbf{C}(u_1, \ldots, u_d) := \mathbf{H}(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))$$

for every $\mathbf{u} \in [0,1]^d$.

Definition 3.2.3. (*Empirical Distribution Function:*) Let 1_A be the indicator function of the set A. We then define the well known empirical distribution for both our multivariate distribution $\mathbf{H}$ and $F_1, \ldots, F_d$ by

$$\mathbf{H}_n(x_1,\ldots,x_d) := \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathbf{1}\{X_{ji}\le x_j\},\ \text{for } \mathbf{x}\in \mathrm{R}^d,$$

and

$$F_{jn}(x_j) := \frac{1}{n}\sum_{i=1}^{n} 1\{X_{ij}\le x_j\} = \mathbf{H}_n(1,\ldots,x_j,\ldots,1),\ \text{for } x_j\in \mathbf{R}.$$

By the strong law of large numbers, the estimators $F_{jn}(t)$ converge almost surely to $F_i(t)$ as n$\rightarrow\infty$, for every value of t. Furthermore, the Glivenko-Cantelli[2] theorem gives an even stronger result, showing that in fact the convergence happens uniformly over t. These results can easily be extended to the multivariate case and give us a simple way to define both the inverse distribution functions F_i^{-1}, and the copula function $C(\mathbf{u})$ as follows

$$F_{in}^{-1}(u_i) = \inf\{t_i\in \mathrm{R} : F_{in}(t_i)\ge u_i\},\ \text{for } u_i\in[0,1],$$

and

$$\mathbf{C}_n(u_1,\ldots,u_d) := \mathbf{H}_n(F_{1n}^{-1}(u_1),\ldots,F_{dn}^{-1}(u_d)),\ \text{for } \mathbf{u}\in[0,1]^d.$$

Set $U_{ji} := F_j(X_{ji})$ for $i=1,\ldots,n$, and $j=1,\ldots,d$. Then for each $n\ge 1$, $0\le u_j\le 1$ and $1\le j\le d$, we have

$$\mathbf{C}_n(u_1,\ldots,u_d) := \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} 1\{U_{ji}\le u_j\},$$

which leads us to the more familiar empirical copula

$$\mathbf{C}(\frac{k_1}{n},\ldots,\frac{k_d}{n}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} 1\{X_{ij}\le x_{(k_j)}\}.$$

To complete our extension we will simply make use of the Riemann sums definition for integrals, first in the two dimensional case, to showcase how Nelson [18], obtained the bivariate empirical Schweizer and Wolff sigma, and then following the same procedure we will extend this to the multivariate case.

Recall the formula for Schweizer and Wolff's sigma is

$$\sigma_{X,Y} = 12\int_0^1\int_0^1 |C(u,v) - uv|dudv \approx$$

$$\approx 12\sum_{i=1}^{n}\sum_{j=1}^{n} |C(u,v) - uv|\Delta u\Delta v,$$

where we take $\Delta u = \Delta v = \frac{1}{n}$ as the widths of the Riemann rectangles. Furthermore, notice that u and v are then simply equally spaced points taking values along the edges of the $[0,1]\times[0,1]$ square. So we can replace them by $\frac{i}{n}$, and $\frac{j}{n}$, respectively. Our Riemann sum representation of the measure then becomes

$$\frac{12}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|C(u,v) - \frac{i}{n}\frac{j}{n}\right|.$$

Clearly, this converges to our integral, as $n \to \infty$, by the properties of Riemann integration. As a final step all we need do is replace $C(u,v)$ with the empiric copula, and we have our empirical Schweizer and Wolff's sigma as follows

$$\frac{12}{n^2-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|C\left(\frac{i}{n},\frac{j}{n}\right) - \frac{i}{n}\frac{j}{n}\right|.$$

Note that we changed the fraction in front of the double sum from $\frac{12}{n^2}$ to $\frac{12}{n^2-1}$. This does not change the limit behaviour of our estimate as $n \to \infty$, but it does make it match the Spearman's rank rho sample statistic which can be calculated from the empiricial copula as follows

$$\frac{12}{n^2-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[C\left(\frac{i}{n},\frac{j}{n}\right) - \frac{i}{n}\frac{j}{n}\right].$$

Following the same process, using Riemann sums for the multidimensional formula for Schweizer and Wolff sigma, we also obtain the d-dimensional empirical formula:

$$\frac{2^d(d+1)}{(2^d-(d+1))(n^d-1)} \sum_{k_1=1}^{n} \cdots \sum_{k_d=1}^{n} \left| C(\frac{k_1}{n} \cdots \frac{k_d}{n}) - \frac{k_1}{n} \cdots \frac{k_d}{n} \right|.$$

Notice that the coefficient in front of the d sums, is simply the coefficient of the d-dimensional Schweizer and Wolff's sigma, $\frac{2^d(d+1)}{2^d-(d+1)}$ obtained in section 3.2.1 above, divided by $n^d - 1$. This is simply $\frac{12}{n^2-1}$ in the two-dimensional case, and in both cases represents the normalization constant used in the theoretical Schweizer and Wolff's sigma calculation, to ensure the measure stays less than or equal to one, divided by the $n^d - 1$ factor born out of the Riemann sum approximation from this section.

4 Application to Specific Bivariate Distribution

In this section we will apply our dependency measures to three bivariate distributions. In particular, we will take a look at the Bivariate Gaussian, a Farlie-Gumble-Morgenstern Exponential Distribution, and finally, a non-stable bivariate distribution with stable marginals. The latter in particular is meant to analyze how each measure behaves when applied with data from distributions without first or second moments.

4.1 Building Bivariate Distributions

In this section we want to lay the groundwork for building specific bivariate and multivariate distribution functions which we will then use in order to calculate explicit dependence measures using the various methods at our disposal in order to be able to compare them. We will start by looking at specific distribution functions and then try to generalize our calculations for specific classes of distribution functions.

In order to do this we will use a technique developed by de la Pena, Ibragimov and Sharakhmetov [4], and extended by Komelji and Perman [14], which allows us to generate the density, distribution and characteristic functions as well as copulas of given classes of multivariate distribution functions. Below we will recall the main results outlined in these papers.

Theorem 4.1.1. [4] *A function* $F : \mathbf{R}^n \to [0, 1]$ *is a joint cdf with one-dimensional marginal cdf's* $F_k(x_k), x_k \in \mathbf{R}, k = 1, \ldots, n$, *absolutely continuous with respect to the product of marginal cdf's* $\prod_{k=1}^n F_k(x_k)$, *if and only if there exist functions* $g_{i_1,\ldots,i_c} : \mathbf{R}^c \to \mathbf{R}$, $1 \leq i_1 < \ldots < i_c \leq n$, $c = 2, \ldots, n$, *satisfying conditions*

A1 *(integrability):*

$$E|g_{i_1,\ldots,i_c}(\xi_{i_1}, \ldots, \xi_{i_c})| < \infty;$$

A2*(degeneracy):*

$$E(g_{i_1,\dots,i_c}(\xi_{i_1},\dots,\xi_{i_{k-1}},\xi_{i_k},\xi_{i_{k+1}},\dots,\xi_{i_c})|\xi_{i_1},\dots,\xi_{i_{k-1}},\xi_{i_{k+1}},\dots,\xi_{i_c}) =$$

$$\int_{-\infty}^{\infty} g_{i_1,\dots,i_c}(\xi_{i_1},\dots,\xi_{i_{k-1}},x_{i_k},\xi_{i_{k+1}},\dots,\xi_{i_c})dF_{i_k}(x_{i_k}) = 0, \text{ (a.s.)},$$

$1 \le i_1 < \dots < i_c \le n$, $k = 1, 2, \dots, c$, $c = 2, \dots, n$,

A3 *(positive definiteness):*

$$U_n(\xi_1,\dots,\xi_n) \equiv \sum_{c=2}^{n} \sum_{1\le i_1<\dots<i_c\le n} g_{i_1},\dots,{i_c}(\xi_{i_1},\dots,\xi_{i_c}) \ge -1 \text{ (a.s.)},$$

and such that the following representation holds for F:

$$F(x_1,\dots,x_n) = \int_{-\infty}^{x_1}\dots\int_{-\infty}^{x_n}(1 + U_n(t_1,\dots,t_n))\prod_{i=1}^{n} dF_i(t_i).$$

Moreover, $g_{i_1,\dots,i_c}(\xi_{i_1},\dots,\xi_{i_c}) = f_{i_1,\dots,i_c}(\xi_{i_1},\dots,\xi_{i_c})(a.s.), 1 \le i_1 < \dots < i_c \le n, c = 2,\dots,n,$ *where*

$$f_{i_1,\dots,i_c}(x_{i_1},\dots,x_{i_c}) = \sum_{k=2}^{c}(-1)^{c-k} \sum_{1\le j_1<\dots<j_k\in\{i_1,\dots,i_c\}} \left(\frac{\partial F(x_{j_1},\dots,x_{j_k})}{\partial F_{j_1}\dots\partial F_{j_k}} - 1\right).$$

Here we take $g_{i_1,\dots,i_k}$ as simply defined as $\frac{\partial F(x_{j_1},\dots,x_{j_k})}{\partial F_{j_1}\dots\partial F_{j_k}}$. It is equal to 1 if at least one $x_{j_1},\dots,x_{j_k}$ is not a point of increase of the corresponding $F_{j_1},\dots,F_{j_k}$.

Using these basic properties, de la Pena et. al. were able to generalize representations for multivariate density and distribution functions as well as the corresponding copulas. Komelji and Perman then extended this to include the characteristic function with the added condition that the function $g_{i_1,\dots,i_k} = g_{i_1}\cdot\ldots\cdot g_{i_k}$. This is a somewhat limiting condition, but we can at least use it in the case of the Farlie-Gumble-Morgenstern family of multivariate distributions.

Theorem 4.1.2. [14] *For $i = 1, \dots, n$ let $g_i : [0,1] \to \mathbf{R}$ be continuous functions which are not identically zero. Assume $\int_0^1 g_i(t)dt = 0$. Let $G_i(t) = \int_0^t g_i(u)du$. Then one can define the copula*

$$C(u_1, \dots, u_n) = \prod_{i=1}^{n} u_i + \sum_{1 \le i < j \le n} \theta_{ij} G_i(u_i) G_j(u_j) \prod_{k=1, k \notin (i,j)}^{n} u_k,$$

and for absolutely continuous $X_1, \dots, X_n$, with cdf's $F_{X_1}, \dots, F_{X_n}$,

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \cdot \left(1 + \sum_{1 \le i < j \le n} \theta_{ij} g_i(F_{X_i}(x_i)) g_j(F_{X_j}(x_j))\right),$$

is a joint pdf which belongs to the absolutely continuous n-dimensional cdf

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) + \sum_{1 \le i < j \le n} \theta_{ij} G_i(F_{X_i}(x_i)) G_j(F_{X_j}(x_j)) \prod_{k=1, k \notin (i,j)}^{n} F_{X_k}(x_k)$$

with one-dimensional marginal cdf's $F_{X_1}, \dots, F_{X_n}$.

The above theorem, stated in Komelji and Perman's paper, is based entirely on results from de la Pena's work, with the added assumption and restriction of generality that $g_{i_1,\dots,i_k} = g_{i_1} \cdot \dots \cdot g_{i_k}$. The following theorem contains the extension to characteristic function representation they introduced in their own paper.

Theorem 4.1.3. [14] *Define $f_{\tilde{Z}}(x) = \left(1 - \frac{g(F_Z(x))}{b}\right) f_Z(x)$, where $b = \max_{0 \le x \le 1} g(x)$. Let $X_1, \dots, X_n$ be absolutely continuous random variables with cdf's $F_{X_1}, \dots, F_{X_n}$, and $F_{\tilde{X}_i}$ to be the cdf's of random variables $\tilde{X}_1, \dots, \tilde{X}_n$. Then the following assertions hold:*

(a) *The function $f_X(x_1, \dots, x_n)$ given by*

$$\prod_{i=1}^{n} f_{X_i}(x_i) + \sum_{1 \le i < j \le n} \theta_{ij} b_i b_j x \left(\left(f_{X_i}(x_i) - f_{\tilde{X}_i}(x_i)\right)\left(f_{X_j}(x_j) - f_{\tilde{X}_j}(x_j)\right) \prod_{k=1, k \notin (i,j)}^{n} f_{X_k}(x_k)\right),$$

is the pdf of a joint distribution with marginal densities $f_{X_1}, \dots, f_{X_n}$.

(b) *The function* $\phi_X(t_1, \dots, t_n)$ *given by*

$$\prod_{i=1}^{n} \phi_{X_i}(t_i) + \sum_{1 \le i < j \le n} \theta_{ij} b_i b_j x \left(\left(\phi_{X_i}(t_i) - \phi_{\tilde{X}_i}(t_i)\right) \left(\phi_{X_j}(t_j) - \phi_{\tilde{X}_j}(t_j)\right) \prod_{k=1, k \notin (i,j)}^{n} \phi_{X_k}(t_k) \right)$$

is a characteristic function of a joint distribution with $\phi_{X_1}, \dots, \phi_{X_n}$ *as marginal cf's.*

The groundwork is now laid to apply these theorems to the Farlie-Gumble-Morgenstern family of distributions which we will use as one of the specific distributions we'll examine.

4.2 Bivariate Normal Distribution

The first distribution we will analyze is the Bivariate Normal. This is a well known and well studied distribution which serves as a good standard for comparison. Below, we outline the joint density and characteristic functions, as well as the copula of this distribution:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right],$$

$$\phi(t, s) = \exp\left[i(t\mu_x + s\mu_y) - \frac{1}{2}(\sigma_x^2 t^2 + 2\rho\sigma_x\sigma_y ts + \sigma_y^2 s^2) \right],$$

$$c(u, v) = N_\rho(\Phi_X^{-1}(u), \Phi_Y^{-1}(v)),$$

where Φ denotes the univariate Normal distribution function, and N_ρ denotes the bivariate Normal distribution function.

4.2.1 Szekely's Distance Correlation

We skip straight to calculating our measures of independence since in the case of the bivariate normal distribution the dependence parameter is in fact Pearson's Product-

Moment Correlation coefficient. The calculations below are reproduced from Szekely's paper [30] for the reader's convenience.

Let us define function $F(\rho)$ as follows:

$$F(\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\phi_{X,Y}(t,s) - \phi_X(t)\phi_Y(s)|^2 \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|e^{i(\mu_X t+\mu_Y s)-\frac{1}{2}(\sigma_X^2 t^2+2\rho\sigma_X\sigma_Y ts+\sigma_Y^2 s^2)} - e^{i\mu_X t-\frac{1}{2}\sigma_X^2 t^2} e^{i\mu_Y s-\frac{1}{2}\sigma_Y^2 s^2}\right|^2 \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} \left|e^{i(\mu_X t+\mu_Y s)} e^{-\frac{1}{2}(\sigma_X^2 t^2+2\rho\sigma_X\sigma_Y ts+\sigma_Y^2 s^2)} - e^{i(\mu_X t+\mu_Y s)} e^{-\frac{1}{2}(\sigma_X^2 t^2+\sigma_Y^2 s^2)}\right|^2 \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} \left|e^{i(\mu_X t+\mu_Y s)}\right|^2 \left|e^{-\frac{1}{2}(\sigma_X^2 t^2+\sigma_Y^2 s^2)}\right|^2 \left|e^{-\rho\sigma_X\sigma_Y ts} - 1\right|^2 \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} e^{-\sigma_X^2 t^2-\sigma_Y^2 s^2} \left(1 - 2e^{-\sigma_X\sigma_Y\rho ts} + e^{-2\sigma_X\sigma_Y\rho ts}\right) \frac{dt}{t^2}\frac{ds}{s^2}.$$

Note tha $F(0) = F^{‘}(0) = 0$, so $F(\rho) = \int_0^\rho \int_0^x F^{”}(z)dzdx$. The second derivative of F is

$$F^{”}(z) = \frac{d^2}{dz^2}\int_{R^2} e^{-\sigma_X^2 t^2-\sigma_Y^2 s^2} \left(1 - 2e^{-\sigma_X\sigma_Y zts} + e^{-2\sigma_X\sigma_Y zts}\right) \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} e^{-\sigma_X^2 t^2-\sigma_Y^2 s^2} \left(-2\sigma_X^2\sigma_Y^2 e^{-\sigma_X\sigma_Y zts} + 4\sigma_X^2\sigma_Y^2 e^{-2\sigma_X\sigma_Y zts}\right) dtds.$$

Applying the following change of variables $u = \sigma_X t, v = \sigma_Y s$, we have

$$F^{”}(z) = \frac{d^2}{dz^2}\int_{R^2} e^{-u^2-v^2} \left(-2\sigma_X^2\sigma_Y^2 e^{-zuv} + 4\sigma_X^2\sigma_Y^2 e^{-2zuv}\right) \frac{du}{\sigma_X}\frac{dv}{\sigma_Y}$$

$$= 4\sigma_X\sigma_Y V(z) - 2\sigma_X\sigma_Y V(\frac{z}{2}),$$

where

$$V(z) = \int_{R^2} e^{-u^2-v^2-2zuv} dudv = \frac{\pi}{\sqrt{1-z^2}}.$$

After another change of variables, and using the fact that the eigenvalues of the quadratic form $t^2 + s^2 + 2zts$ are $1 \pm z$, and $\int_{-\infty}^{\infty} e^{-t^2\lambda}dt = (\pi/\lambda)^{1/2}$, we have the following result for $F(\rho)$

$$F(\rho) = \int_0^\rho \int_0^x \left(\frac{4\sigma_X\sigma_Y\pi}{\sqrt{1-z^2}} - \frac{2\sigma_X\sigma_Y\pi}{\sqrt{1-z^2/4}} \right) dzdx$$

$$= 4\sigma_X\sigma_Y\pi \int_0^\rho (\arcsin(x) - \arcsin(x/2))dx$$

$$= 4\sigma_X\sigma_Y\pi(\rho \arcsin \rho + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1).$$

Finally, we can now produce our result for the distance correlation as follows

$$R^2(X,Y) = \frac{F(\rho)}{F(1)} = \frac{\rho \arcsin \rho + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1}{1 + \frac{\pi}{3} - \sqrt{3}}.$$

Thus, we see that R^2, in this case, is a function of ρ only.

4.2.2 Schweizer and Wolff's Sigma

To calculate the Schweizer and Wolff's Sigma for the Bivariate Normal Distribution we will make use of the fact that for a given ρ this distribution is either Positivie Quadrant Dependent or Negative Quadrant Dependent. As such we know that our copula based dependence measure $\sigma_{X,Y}$ is going to be equal to the absolute value of Spearman's Rank Rho. Recall that

$$\rho_{X,Y} = 12 \int_0^1 \int_0^1 (C(u,v,\rho) - uv)\, dudv$$

$$= 12 \int_0^1 \int_0^1 C(u,v,\rho)\, dudv - 3$$

$$= \frac{6}{\pi} \arcsin \frac{\rho}{2}.$$

The proof can be found in Embrechts [7]. We can then use this to calculate the Schweizer and Wolff sigma which is given by:

$$\sigma_{X,Y} = 12 \int\int_{I^2} |C(u,v) - uv| \, dudv.$$

Clearly, if X and Y are PQD (Positive Quadrant Dependant), then $\sigma_{X,Y} = \rho_{X,Y}$, and if they are NQD, $\sigma_{X,Y} = -\rho_{X,Y}$.

Thus,

$$\sigma_{X,Y} = \left| \frac{6}{\pi} \arcsin \frac{\rho}{2} \right|.$$

4.3 Farlie-Gumble-Morgenstern Family of Distributions

Recall the bivariate Farlie Gumble Morgenstern (FGM) distribution has the following distribution function with given marginal distributions $F_1(x_1)$, and $F_2(x_2)$,

$$F_\theta(x_1, x_2) = F_1(x_1)F_2(x_2)(1 + \theta(1 - F_1(x_1))(1 - F_2(x_2)).$$

Using the methods outlined above we quickly obtain the following identities which we then used to define our density and characteristic functions as well as the copulas for this family of bivariate distributions:

$$\frac{dF_\theta(x_1, x_2)}{dF_1(x_1)dF_2(x_2)} = \theta(1 - 2t_1)(1 - 2t_2) \qquad \Rightarrow g_i(t_i) = 1 - 2t_i$$

$$G_i(x) = \frac{dg_i(x)}{dx} = x - x^2 \qquad \Rightarrow h_i(x) = x - \frac{G_i(x)}{\max_{0 \leq x \leq 1} g_i(x)} = x^2.$$

We thus have the following general bivariate FGM density and characteristic functions and copula in terms of the underlying marginals:

$$f_\theta(x_1, x_2) = f_1(x_1)f_2(x_2) \left[1 + \theta(1 - 2F_1(x_1))(1 - 2F_2(x_2))\right],$$

$$\phi_\theta(x_1, x_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2) \left[1 + \theta \left(\phi_{X_1}(t_1) - \phi_{\tilde{X}_1}(t_1)\right) \left(\phi_{X_2}(t_2) - \phi_{\tilde{X}_2}(t_2)\right)\right],$$

$$c_\theta(u,v) \quad = uv(1+\theta(1-u)(1-v)),$$

where, $\phi_{\tilde{X}_i}(t_i) = E_{\tilde{X}} e^{it_1 x}$, for density $f_{\tilde{X}_i}(x_i) = [1-g_i(F_i(x))]f_i(x_i)$.

We will set about attempting to write Pearson's Rank Moment Correlation, Distance Correlation, and Schweizer and Wolff's sigma in terms of θ, with the aid of the above generic bivariate FGM equations.

4.3.1 Pearson's Rank Moment Correlation

Recall the following about ρ:

$$\rho = \frac{E(X_1X_2) - E(X_1)E(X_2)}{\sigma_{x_1}\sigma_{x_2}},$$
$$E(X_1X_2) = \int_{-\infty}^{\infty} x_2 E(X_1|X_2 = x_2) f_2(x_2) dx_2,$$

where

$$f(x_1|x_2) = f(x,y)/f_2(y) = f_1(x_1)\left[1+\theta(1-2F_1(x_1))(1-2F_2(x_2))\right].$$

Thus,

$$E(X_1|X_2 = x_2) = \int_{-\infty}^{\infty} x_1 f_1(x_1)\left[1+\theta(1-2F_1(x_1))(1-2F_2(x_2))\right]dx_1$$
$$= E(X_1) + \theta(1-2F_2(x_2))\int_{-\infty}^{\infty} x_1 f_1(x_1)(1-2F_1(x_1))dx_1$$
$$= E(X_1) + \theta(1-2F_2(x_2))(E(X_1) - 2E(X_1F_1(x_2))).$$

Therefore,

$$E(X_1,X_2) = \int_{-\infty}^{\infty} x_2[E(X_1) + \theta(1-2F_2(x_2))(E(X_1) - 2E(X_1F_1(x_2)))]f_2(x_2)dx_2$$

$$= E(X_1)E(X_2) + \theta[E(X_1) - 2E(X_1F_1(x_1))] \int_{-\infty}^{\infty} x_2(1 - 2F_2(x_2))f_2(x_2)dx_2$$

$$= E(X_1)E(X_2) + \theta[E(X_1) - 2E[X_1F_1(x_1))][E(X_2) - 2E(X_2F_2(x_2))].$$

Finally, we have the following expression for Pearson's Rank Moment Correlation of a bivariate FGM distribution:

$$\rho = \frac{E(X_1)E(X_2) + \theta[E(X_1) - 2E[X_1F_1(x_1))][E(X_2) - 2E(X_2F_2(x_2))] - E(X_1)E(X_2)}{\sigma_{x_1}\sigma_{x_2}}$$

$$= \frac{\theta[E(X_1) - 2E[X_1F_1(x_1))][E(X_2) - 2E(X_2F_2(x_2))]}{\sigma_{x_1}\sigma_{x_2}}.$$

4.3.2 Distance Correlation

Applying Szekely's distance correlation equation to the bivariate characteristic function we have developed using de la Pena's and Komelji's method, we have the following results:

$$V^2(X,Y) = \frac{1}{\pi^2}\int_{R^2} \frac{|\phi_{X,Y}(t,s) - \phi_X(t)\phi_Y(s)|^2}{t^2s^2} dtds$$

$$= \frac{1}{\pi^2}\int_{R^2} \frac{|\phi_X(t)\phi_Y(s) + \theta\left(\phi_X(t) - \phi_{\tilde{X}}(t)\right)\left(\phi_Y(s) - \phi_{\tilde{Y}}(s)\right) - \phi_X(t)\phi_Y(s)|^2}{t^2s^2} dtds$$

$$= \frac{\theta}{\pi^2}\int_{R^2} \frac{|\left(\phi_X(t) - \phi_{\tilde{X}}(t)\right)\left(\phi_Y(s) - \phi_{\tilde{Y}}(s)\right)|^2}{t^2s^2} dtds,$$

and

$$V^2(X) = \frac{1}{\pi^2}\int_{R^2} \frac{|\phi_X(t+s) - \phi_X(t)\phi_X(s)|^2}{t^2s^2} dtds$$

$$V^2(Y) = \frac{1}{\pi^2}\int_{R^2} \frac{|\phi_Y(t+s) - \phi_Y(t)\phi_Y(s)|^2}{t^2s^2} dtds.$$

It then follows that Szekely's Distance Correlation is given by

$$R^2(X,Y) = \frac{V^2(X,Y)}{\sqrt{V^2(X)V^2(Y)}},$$

where $\phi_{\tilde{X}}(t)$ and $\phi_{\tilde{Y}}(s)$ are the characteristic functions of random variables $\tilde{X}$ and $\tilde{Y}$ with density functions $f_{\tilde{X}}(x) = [1 - g(F_X(x))]f_X(x) = 2f_X(x)F_X(x)$ and $f_{\tilde{Y}}(y) = 2f_Y(y)F_Y(y)$. Finally, we must note here that Szekely defines $|f(x)|^2$, where $f(x)$ is a complex function as $f(x)\tilde{f}(x)$, where $\tilde{f}(x)$ is the complex conjugate of $f(x)$.

4.3.3 Schweizer and Wolff's Sigma

Using the classic Schweizer and Wolff's sigma expression, as well as the copula defined for the FGM family of distributions, we have the following expression:

$$\begin{aligned}
\sigma_{X,Y} &= 12 \int_{I^2} |C_\theta(u,v) - uv| dudv \\
&= 12 \int_{I^2} |uv(1 + \theta(1-u)(1-v)) - uv| dudv \\
&= 12 \int_{I^2} |uv + \theta uv(1-u)(1-v) - uv| dudv \\
&= 12 \int_{I^2} |\theta uv(1-u)(1-v)| dudv \\
&= 12 \int_{I^2} |\theta uv(1 - u - v + uv)| dudv \\
&= 12 \int_{I^2} |\theta(uv - u^2v - uv^2 + u^2v^2)| dudv \\
&= \frac{12\theta}{36} = \frac{\theta}{3}.
\end{aligned}$$

It is worthwhile to notice here that for the bivariate FGM the Schweizer and Wolff's Sigma is equal to Spearman's Rank Rho, a fact that can be easily proved using the copula version of Spearman's Rank Rho.,

4.4 Bivariate Exponential FGM

We can now apply these results to a specific bivariate FGM distribution with exponential marginals. We will have the same Schweizer and Wolff's Sigma, since this dependence measure is based purely on Copulas which are independent of the underlying marginals. So we only need to calculate Pearson's Rank Moment Correlation and Szekely's Distance Correlation in terms of theta. Recall the pdf of the Exponential Distribution is $f(x) = \lambda e^{-\lambda x}$.

4.4.1 Bivariate Exponential FGM Correlation

Using previous results for the correlation of bivariate FGM distributions we can calculate the correlation in terms of θ for the bivariate exponential distribution as follows.

$$\rho = \frac{\theta[E(X_1) - 2E[X_1F_1(X_1))][E(X_2) - 2E(X_2F_2(X_2))]}{\sigma_{X_1}\sigma_{X_2}}.$$

To calculate this, we must evaluate the value of $E[x_1F_1(x_1)]$:

$$E[XF(X)] = E[X(1-e^{-\lambda X})] = \int_0^\infty \left(x - xe^{-\lambda x}\right) f(x)dx = E(X) - \lambda \int_0^\infty xe^{-2\lambda x}dx =$$

$$= \frac{1}{\lambda} - \lambda\left[x\frac{-1}{2\lambda}e^{-2\lambda x}\Big|_0^\infty - \int_0^\infty \frac{-1}{2\lambda}e^{-2\lambda x}dx\right] = \frac{1}{\lambda} - \lambda\left[0 - \frac{1}{4\lambda^2}e^{-2\lambda x}\Big|_0^\infty\right] = \frac{1}{\lambda} - \frac{1}{4\lambda}.$$

Thus we have $E[X_1F_1(X_1)] = E[X_2F_2(X_2)] = \frac{3}{4\lambda}$.

Then the Pearson Rank Moment Correlation for the Bivariate Exponential FGM is

$$\rho = \frac{\theta[\frac{1}{\lambda_X} - 2\frac{3}{4\lambda_X}][\frac{1}{\lambda_Y} - 2\frac{3}{4\lambda_Y}]}{\frac{1}{\lambda_X\lambda_Y}} = \theta\lambda_X\lambda_Y\frac{-1}{2\lambda_X}\frac{-1}{2\lambda_Y} = \frac{\theta}{4}.$$

4.4.2 Bivariate Exponential FGM Distance Correlation

The characteristic function of the Bivariate Exponential FGM Distribution is calculated using Komelji's work [14], and reproduced below:

$$\phi_\theta(t,s) = \frac{\lambda_X\lambda_Y}{(\lambda_X - it)(\lambda_Y - is)}\left(1 - \frac{\theta ts}{(2\lambda_X - it)(2\lambda_Y - is)}\right).$$

Now we will first calculate the the distance Covariance defined in Szekely's paper as follows:

$$\begin{aligned}
V^2(X,Y) &= \frac{1}{\pi^2}\int_{R^2} \frac{|\phi_{X,Y}(t,s) - \phi_X(t)\phi_Y(s)|^2}{t^2s^2} dtds \\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\left|\frac{\lambda_X\lambda_Y}{(\lambda_X - it)(\lambda_Y - is)}\left(1 - \frac{\theta ts}{(2\lambda_X - it)(2\lambda_Y - is)}\right) - \frac{\lambda_X\lambda_Y}{(\lambda_X - it)(\lambda_Y - is)}\right|^2}{t^2s^2} dtds \\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\left|\frac{-\theta\lambda_X\lambda_Y ts}{(2\lambda_X - it)(\lambda_X - it)(2\lambda_Y - is)(\lambda_Y - is)}\right|^2}{t^2s^2} dtds \\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\theta^2\lambda_X^2\lambda_Y^2}{|(2\lambda_X - it)(\lambda_X - it)(2\lambda_Y - is)(\lambda_Y - is)|^2} dtds \\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\theta^2\lambda_X^2\lambda_Y^2}{|(2\lambda_X^2 - 3\lambda_X it - t^2)(2\lambda_Y^2 - 3\lambda Y is - s^2)|^2} dtds \\
&= \frac{\theta^2\lambda_X^2\lambda_Y^2}{\pi^2}\frac{\pi^2}{36\lambda_X^3\lambda_Y^3} = \frac{\theta^2}{36\lambda_X\lambda_Y}.
\end{aligned}$$

To finish up our calculation of Szekely's Distance Correlation, we must now calculate the Distance Variances for X and Y. Recall that

$$V^2(X,X) = \frac{1}{\pi^2}\int_{R^2} \frac{|\phi_{X,X}(t,s) - \phi_X(t)\phi_X(s)|^2}{t^2s^2} dtds.$$

where $\phi_{X,X}(t,s)$ represents $E_X e^{i(t+s)}$; thus,

$$\begin{aligned}
V^2(X,X) &= \frac{1}{\pi^2}\int_{R^2} \frac{|\frac{\lambda}{\lambda_- i(t+s)} - \frac{\lambda^2}{(\lambda - it)(\lambda - is)}|^2}{t^2s^2} dtds \\
&= \frac{1}{\pi^2}\int_{R^2} \frac{\left|\frac{\lambda(\lambda - it)(\lambda - is) - \lambda^2(\lambda - i(t+s))}{(\lambda - i(t+s))(\lambda - it)(\lambda - is)}\right|^2}{t^2s^2} dtds \\
&= \frac{1}{\pi^2}\int_{R^2} \frac{\left|\frac{\lambda^3 - \lambda^2 it - \lambda^2 is - \lambda ts) - \lambda^3 + \lambda^2 it + \lambda^2 is}{(\lambda - i(t+s))(\lambda - it)(\lambda - is)}\right|^2}{t^2s^2} dtds
\end{aligned}$$

$$= \frac{1}{\pi^2} \int_{R^2} \frac{\left| \frac{-\lambda ts}{(\lambda - i(t+s))(\lambda - it)(\lambda - is)} \right|^2}{t^2 s^2} dtds$$

$$= \frac{\lambda^2}{\pi^2} \int_{R^2} \left| \frac{1}{(\lambda - i(t+s))(\lambda - it)(\lambda - is)} \right|^2 dtds$$

$$= \frac{\lambda^2}{\pi^2} \frac{\pi^2}{3\lambda^4} = \frac{1}{3\lambda^2}.$$

We can now calculate the Distance Correlation for the FGM Bivariate Exponential distribution to be

$$R^2(X,Y) = \frac{V^2(X,Y)}{\sqrt{V^2(X,X)V^2(Y,Y)}} = \frac{\theta^2/36\lambda_X\lambda_Y}{1/3\lambda_X\lambda_Y} = \frac{\theta}{\sqrt{12}}.$$

4.5 Comparing the Two Measures of Dependence across the Two Distributions of Interest

Using the results developed above we can see how the two dependence measures perform when applied to distinct bivariate distributions, as well as how they compare with the classic Pearson's product moment correlation.

4.5.1 Inequalities for the FGM Exponential distribution

First notice that in the FGM Exponential case we have

$$\sigma^2_{X,Y} \leq R^2(X,Y) \leq \rho^2.$$

This is trivial from the calculations for each measure of association in the FGM Exponential case.

4.5.2 Inequalities for the Bivariate Normal Distribution

Now we turn to the case of the bivariate normal distribution, where we will show that

$$R^2(X,Y) \leq \sigma^2_{X,Y} \leq \rho^2.$$

The proof of this is not trivial, and we will introduce three propositions below to prove these inequalities.

Proposition 4.5.1. *Given two random variables X and Y with a joint normal distribution function and given correlation ρ, then*

$$R^2(X,Y) \leq \rho^2,$$

where $R^2(X,Y)$ is Szekely's distance correlation.

Proof. Reproduced from Szekely's paper [30]. Let X and Y be Standard Normal random variables, with $Cov(X,Y) = \rho(X,Y) = \rho$. Define

$$F(\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f_{X,Y}(t,s) - f_X(t)f_Y(s)|^2 \frac{dt}{t^2}\frac{ds}{s^2}.$$

Then dCov, $V^2(X,Y) = F(\rho)/c_1^2 = F(\rho)/\pi^2$ and

$$R^2(X,Y) = \frac{V^2(X,Y)}{\sqrt{V^2(X,X)V^2(Y,Y)}} = \frac{F(\rho)}{F(1)},$$

where

$$F(\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |e^{-(t^2+s^2)/2-\rho ts} - e^{-t^2/2}e^{-s^2/2}|^2 \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} e^{-t^2-s^2}(1 - 2e^{-\rho ts} + e^{-2\rho ts})\frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} e^{-t^2-s^2} \sum_{n=2}^{\infty} \frac{2^n-2}{n!}(-\rho ts)^n \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \int_{R^2} e^{-t^2-s^2} \sum_{k=1}^{\infty} \frac{2^{2k}-2}{(2k)!}(-\rho ts)^{2k} \frac{dt}{t^2}\frac{ds}{s^2}$$

$$= \rho^2 \left[\sum_{k=1}^{\infty} \frac{2^{2k}-2}{(2k)!}\rho^{2(k-1)} \int_{R^2} e^{-t^2-s^2}(ts)^{2(k-1)} dtds \right].$$

Thus, $F(\rho) = \rho^2 G(\rho)$, where $G(\rho)$ is a sum with all nonnegative terms. Clearly, as ρ increases, so does $G(\rho)$, and $G(\rho) \leq G(1)$. Therefore,

$$R^2(X,Y) = \frac{F(\rho)}{F(1)} = \rho^2 \frac{G(\rho)}{G(1)} \leq \rho^2.$$

□

Proposition 4.5.2. *Let X and Y be two random vectors with joint normal distribution, and a given correlation ρ. Then the following inequality holds*

$$\sigma^2_{X,Y} \leq \rho^2,$$

where $\sigma^2_{X,Y}$ represents the Scweizer and Wolff's sigma for the given bivariate distribution function.

Proof. Setting $\rho \in [0,1]$, we will focus on the positive values of correlation for now. Notice that on $x \in [0,1], \arcsin(x)$ has the second derivative,

$$\frac{d^2 \arcsin(x)}{dx^2} = \frac{d\frac{1}{\sqrt{1-x^2}}}{dx} = \frac{x}{(\sqrt{1-x^2})^3}.$$

Clearly on $x \in [0,1]$, this second derivative is positive, and thus, $\arcsin(x)$ is a convex function on this interval. Furthermore, notice that $\arcsin(0) = 0$ and $\arcsin \frac{1}{2} = \frac{\pi}{6}\rho$. Thus, on $\rho \in [0,1]$, we have the line $\frac{\pi}{6}\rho$ always greater then $\arcsin \frac{\rho}{2}$. Thus,

$$\arcsin\frac{\rho}{2} \leq \frac{\pi}{6}\rho \Rightarrow \left|\frac{6}{\pi}\arcsin\frac{\rho}{2}\right| \leq \frac{6}{\pi}\frac{\pi}{6}\rho = \rho.$$

Thus, for positive values of $\rho, \sigma_{X,Y} \leq \rho$. For negative values of ρ, we use the following identity $\arcsin(-x) = -\arcsin(x)$, and we have

$$\frac{6}{\pi}\arcsin -\frac{\rho}{2} = -\frac{6}{\pi}\arcsin\frac{\rho}{6}.$$

Therefore, for $\rho \in [-1,0], \sigma_{X,Y} = \left|\frac{6}{\pi}\arcsin\frac{\rho}{2}\right| = \frac{6}{\pi}\arcsin\frac{\rho}{2}$ if we now take $\rho \in [0,1]$. Thus clearly for $\rho \in [-1,1]$,

$$\left|\frac{6}{\pi}\arcsin\frac{\rho}{2}\right|^2 \leq \rho^2.$$

□

Proposition 4.5.3. *Given two random variables X and Y with joint normal distribution function, then for $R^2(X,Y)$, Szekelys distance correlation, and $\sigma^2_{X,Y}$, the Schweizer and Wolff's sigma, the following inequality holds*

$$R^2(X,Y) \leq \sigma^2_{X,Y}.$$

Proof. We will show this by first noting that for $\rho = 0$ or $\rho = 1, \sigma^2_{X,Y} - R^2(X,Y) = 0$. Thus, if we can show the function $\sigma^2_{X,Y} - R^2(X,Y)$ has only the two minima at $\rho = 0$ and $\rho = 1$ on $\rho \in [0,1]$, then clearly the function is positive on that domain and we have $R^2(X,Y) \leq \sigma^2_{X,Y}$. Indeed,

$$\sigma^2_{X,Y} - R^2(X,Y) =$$

$$= \left(\frac{6}{\pi}\arcsin\frac{\rho}{2}\right)^2 - \frac{\rho\arcsin\rho + \sqrt{1-\rho^2} - \rho\arcsin\rho/2 - \sqrt{4-\rho^2} + 1}{1 + \pi/3 - \sqrt{3}}$$

$$= \frac{(1+\pi/3-\sqrt{3})\frac{36}{\pi^2}\arcsin^2\frac{\rho}{2} - \rho(\arcsin\rho - \arcsin\frac{\rho}{2}) - \sqrt{1-\rho^2} + \sqrt{4-\rho^2} - 1}{1+\pi/3-\sqrt{3}}.$$

Clearly, whether this is a positive or negative function of ρ is determined by the numerator of this fraction,

$$(1+\pi/3-\sqrt{3})\frac{36}{\pi^2}\arcsin^2\frac{\rho}{2} - \rho(\arcsin\rho - \arcsin\frac{\rho}{2}) - \sqrt{1-\rho^2} + \sqrt{4-\rho^2} - 1 =$$

$$= \frac{36(1-\sqrt{3})+12\pi}{\pi^2}\arcsin^2\frac{\rho}{2} - \rho(\arcsin\rho - \arcsin\frac{\rho}{2}) - \sqrt{1-\rho^2} + \sqrt{4-\rho^2} - 1.$$

Let this last simplification of the numerator be called $F(\rho)$. We will now take the derivative of this function to find its maxima and minima and see how it behaves on the interval [0, 1],

$$\frac{dF(\rho)}{d\rho} = \frac{72(1-\sqrt{3})+24\pi}{\pi^2}\arcsin\frac{\rho}{2}\frac{1}{\sqrt{4-\rho^2}} - \arcsin\rho + \arcsin\frac{\rho}{2} -$$

$$-\rho\left(\frac{1}{\sqrt{1-\rho^2}} - \frac{1}{\sqrt{4-\rho^2}}\right) - \frac{1}{2}\frac{1}{\sqrt{1-\rho^2}}(-2\rho) + \frac{1}{2}\frac{1}{\sqrt{4-\rho^2}}(-2\rho) =$$

$$= \frac{72(1-\sqrt{3})+24\pi}{\pi^2\sqrt{4-\rho^2}}\arcsin\frac{\rho}{2} - \arcsin\rho + \arcsin\frac{\rho}{2}.$$

Setting this to 0 and solving, we get the roots of this function to be $\rho = 0$ and $\rho \approx \pm 0.8$. Upon closer inspection, we notice that $F'(\rho)$ is positive before $\rho \approx \pm 0.8$ and negative right after. So $\rho \approx \pm 0.8$ are maxima, and similarly, we can see that $\rho = 0$ is a minima. Thus $F(\rho)$ has a minima at $\rho = 0$ and $F(0) = 0$. Furthermore $F(\pm 1) = 0$, and the function also has maxima at $\rho \approx \pm 0.8$. Thus the function starts at 0 when $\rho = -1$ increases until $\rho \approx -0.8$ reaching a maxima, decreasing once more until $\rho = 0$ where it has a minima and increases once more until $\rho \approx 0.8$ where it has another maxima, and then decreases once more hitting the value 0 at $\rho = 1$. Thus its obvious that $F(\rho) \geq 0$ on the whole interval on which it is defined, $\rho \in [-1, 1]$.

Therefore,

$$\sigma_{X,Y}^2 \geq R^2(X,Y), \text{ for } \rho \in [-1,1].$$

□

Taking these three propositions together gives us the desired inequality. In both cases, we notice that both measures of dependence are consistently lower then Pearson's product moment, but there is no general statement of inequality we can make between them that extends to all joint distribution functions.

We conclude this section by providing a graph of the three measures of association across the two joint distribution functions we have studied thus far. This is useful because it not only gives a visual representation of the inequalities above, but it also shows that the values stick very close to each other across the full spectrum of association, ranging from independence to full dependence.

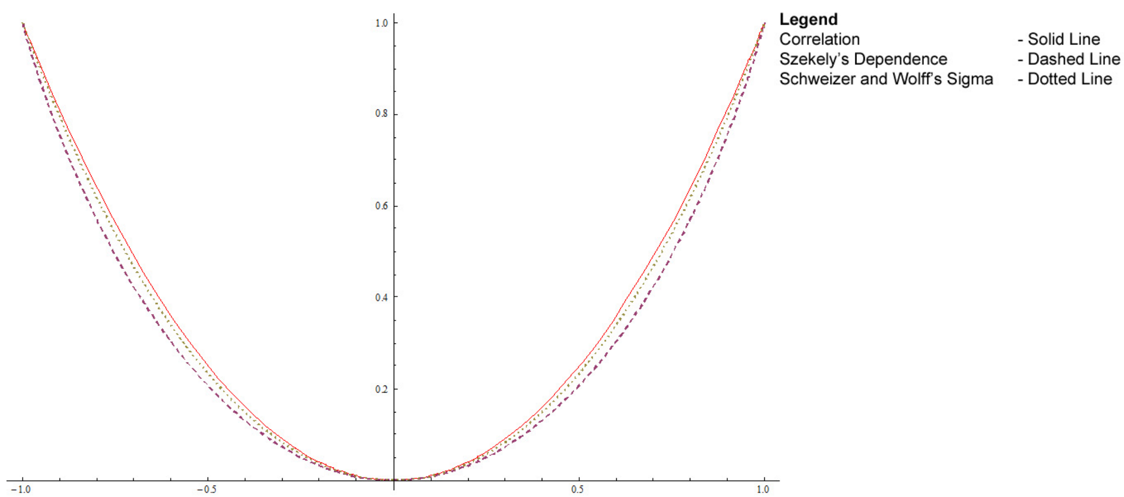

Figure 1: Three Measures of Association for the Bivariate Normal Distribution

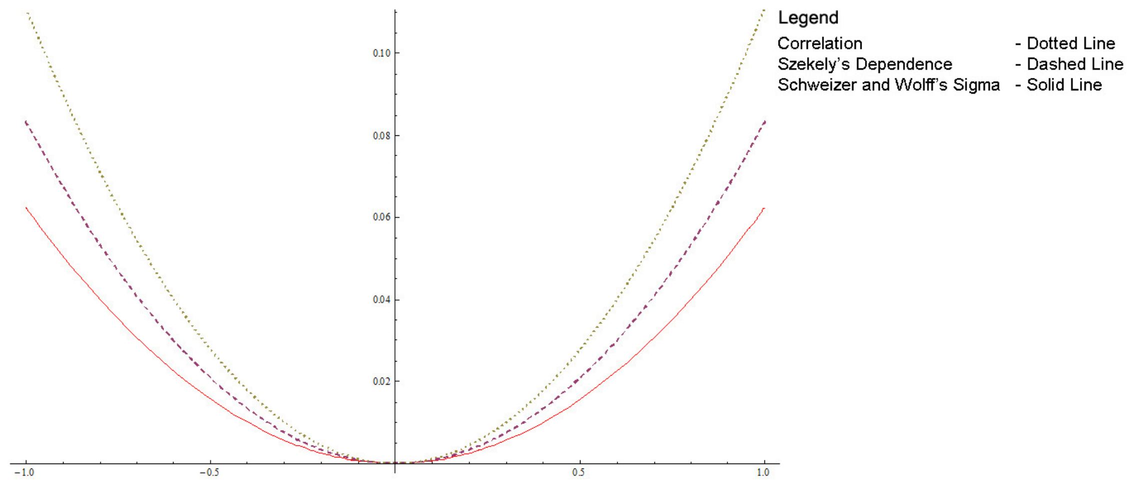

Figure 2: Three Measures of Association for the FGM Bivaraite Exponential Distribution

4.6 A non-Stable Bivariate Distribution with Stable Marginals

We conclude this chapter by taking a look at a specific bivariate distribution without first or second moments. This excludes the use of Pearson's product moment correlation and, allows us to see the behaviour of our measures of dependence in such situations. We also suspect that at least one of the measures will perform well, and thus embolden us to try and loosen the finiteness requirements placed on them.

The distribution in question is a bivariate distribution with stable marginals but which is not itself stable, where a distribution is stable if it has the property that a linear combination of two independent copies of the variable has the same distribution, up to location and scale parameters. It was introduced by Marcus [16] in a paper meant to show that unlike bivariate distributions with normal marginals, which are themselves normal, one could have bivariate distributions with stable marginals, without the joint distribution being stable itself.

The original format of it's characteristic function is given bellow in Polar Coordinates:

$$\phi_{X,Y}(r,\theta) = e^{i\alpha r\cos 3\theta - r^{\gamma}}.$$

Converting to cartesian coordinates, using s as the x-axis and t as the y-axis, and for $\frac{1}{2} < \gamma < 1$, we have

$$\phi_{X,Y}(t,s) = e^{i\alpha\sqrt{s^2+t^2}\cos 3(\arctan\frac{s}{t})-(s^2+t^2)^{\gamma/2}}.$$

The resulting marginals are

$$\phi_X(t) = e^{i\alpha t - t^\gamma},$$

and

$$\phi_Y(s) = e^{i\alpha s\cos\frac{3\pi}{2} - s^\gamma} = e^{-s^\gamma}.$$

Converting back to polar coordinates our marginals are then

$$\phi_X(r,\theta) = e^{i\alpha r\cos\theta - r^\gamma\cos^\gamma\theta},$$

and

$$\phi_Y(r,\theta) = e^{-r^\gamma\cos^\gamma\theta}.$$

It is trivial to show that these marginals are in fact stable.

4.6.1 Szekely's Distance Correlation

We know that the D.J. Marcus example does not have first moments, and we now wish to highlight why we insisted on lowering the moment requirements for Szekely's distance correlation. Below we show how using Szekely's distance correlation as defined in [30] would not lead to a converging measure of dependence,

$$\begin{aligned} R(X,Y) &= \frac{1}{c_p c_q}\int_0^{2\pi}\int_0^\infty \frac{|\phi_{X,Y}(t,s) - \phi_X(t)\phi_Y(s)|^2}{|r\cos\theta|^2|r\sin\theta|^2} r dr d\theta \\ &= \frac{1}{\pi^2}\int_0^{2\pi}\int_0^\infty \frac{|\phi_{X,Y}(t,s) - \phi_X(t)\phi_Y(s)|^2}{|r\cos\theta|^2|r\sin\theta|^2} r dr d\theta \end{aligned}$$

$$= \frac{1}{\pi^2}\int_0^{2\pi}\int_0^{\infty}\frac{|e^{i\alpha r\cos 3\theta - r^{\gamma}} - e^{i\alpha r\cos\theta - r^{\gamma}\cos^{\gamma}\theta}e^{-r^{\gamma}\cos^{\gamma}\theta}|^2}{|r\cos\theta|^2|r\sin\theta|^2} r dr d\theta$$

$$= \frac{1}{\pi^2}\int_0^{2\pi}\int_0^{\infty}\frac{|e^{-r^{\gamma}}e^{i\alpha r\cos 3\theta} - e^{-2r^{\gamma}\cos^{\gamma}\theta}e^{i\alpha r\cos\theta}|^2}{|r\cos\theta|^2|r\sin\theta|^2} r dr d\theta.$$

Simplifying the numerator of the fraction inside our integral, and note that,

$$|e^{-r^{\gamma}}e^{i\alpha r\cos 3\theta} - e^{-2r^{\gamma}\cos^{\gamma}\theta}e^{i\alpha r\cos\theta}|^2 =$$

$$= |e^{-r^{\gamma}}[\cos(\alpha r\cos 3\theta) + i\sin(\alpha r\cos 3\theta)] - e^{-2r^{\gamma}\cos^{\gamma}\theta}[\cos(\alpha r\cos\theta) + i\sin(\alpha r\cos\theta)]|^2$$

$$= \Big|\Big[e^{-r^{\gamma}}\cos(\alpha r\cos 3\theta) + e^{-2r^{\gamma}\cos^{\gamma}\theta}\cos(\alpha r\cos\theta)\Big] +$$

$$+ i\Big[e^{-r^{\gamma}}\sin(\alpha r\cos 3\theta) + e^{-2r^{\gamma}\cos^{\gamma}\theta}\sin(\alpha r\cos\theta)\Big]\Big|2$$

$$= \Big[e^{-r^{\gamma}}\cos(\alpha r\cos 3\theta) + e^{-2r^{\gamma}\cos^{\gamma}\theta}\cos(\alpha r\cos\theta)\Big]^2 +$$

$$+ \Big[e^{-r^{\gamma}}\sin(\alpha r\cos 3\theta) + e^{-2r^{\gamma}\cos^{\gamma}\theta}\sin(\alpha r\cos\theta)\Big]^2$$

$$= \Big[e^{-2r^{\gamma}}\cos^2(\alpha r\cos 3\theta) + e^{-4r^{\gamma}\cos^{\gamma}\theta}\cos^2(\alpha r\cos\theta) +$$

$$+ 2e^{-r^{\gamma}(1-2\cos^{\gamma}\theta)}\cos(\alpha r\cos 3\theta)\cos(\alpha r\cos\theta)\Big] +$$

$$+ \Big[e^{-2r^{\gamma}}\sin^2(\alpha r\cos 3\theta) + e^{-4r^{\gamma}\cos^{\gamma}\theta}\sin^2(\alpha r\cos\theta) +$$

$$+ 2e^{-r^{\gamma}(1-2\cos^{\gamma}\theta)}\sin(\alpha r\cos 3\theta)\sin(\alpha r\cos\theta)\Big]$$

$$= e^{-2r^{\gamma}} + e^{-4r^{\gamma}\cos^{\gamma}\theta} + 2e^{-r^{\gamma}(1-2\cos^{\gamma}\theta)}\cos(\alpha r(\cos 3\theta - \cos\theta)).$$

Thus, our Distance Correlation simplifies to:

$$\frac{1}{\pi^2}\int_0^{2\pi}\int_0^{\infty}\frac{e^{-2r^{\gamma}} + e^{-4r^{\gamma}\cos^{\gamma}\theta} + 2e^{-r^{\gamma}(1-2\cos^{\gamma}\theta)}\cos(\alpha r(\cos 3\theta - \cos\theta))}{|r\cos\theta|^2|r\sin\theta|^2} r dr d\theta$$

$$= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{\infty}\frac{e^{-2r^{\gamma}} + e^{-4r^{\gamma}\cos^{\gamma}\theta} + 2e^{-r^{\gamma}(1-2\cos^{\gamma}\theta)}\cos(\alpha r(\cos 3\theta - \cos\theta))}{r^3\sin^2 2\theta} dr d\theta.$$

The question is, do the marginal distributions of our bivariate stable distribution have first moments, which is a requirement for Szekely's distance correlation? Checking the characteristic function's first derivatives,

$$\frac{\delta\phi_X(t)}{\delta t} = \frac{\delta(e^{i\alpha t - t^\gamma})}{\delta t} = i\alpha - \gamma t^{\gamma-1} e^{i\alpha t - t^\gamma}.$$

Observe its divergence as $t \to 0$, because $0 < \gamma < 1$.

Similarly, we inspect the moments of the second marginal distribution,

$$\frac{\delta\phi_Y(s)}{\delta s} = \frac{\delta(e^{-s^\gamma})}{\delta s} = -\gamma s^{\gamma-1} e^{-s^\gamma}.$$

As with the previous marginal distribution this diverges as $s \to 0$, because $0 < \gamma < 1$. However, as we showed in Chapter 3, using $\alpha < \gamma$ in Szekely's distance correlation, does lead to convergence and allows us to confidently apply the measure to the data set generated from the D.J. Marcus distribution.

4.6.2 Analyzing the Dependence Structure

Due to the nature of the D.J. Marcus example, analyzing the dependence structure analytically is exceedingly difficult. Moreover, we wish to include an applied component to our thesis, where we can see actual implementations of two of the dependence measures on simulated data. We can do this by generating a discrete estimate of the cumulative distribution function, using the well known Inversion Formula. We make use of the Framework developed by Shepard [26] which is a variation of the classic approach proposed by Gurland [12] and refined by Gil-Pelaez [11]. Shepard's version is used primarily because he gives a very nice extension of the Inversion Formula to the multivariate case, but also because in his follow up paper [27] on the subject he gives a good application of Numerical Methods for implementing the Inversion Formula.

Here we will reproduce some of the main results in Shepard's papers, which we used to obtain our CDF estimate for the D.J. Marcus distribution. Let f and ϕ be the density and characteristic functions, respectively, then the application of the Fourier inversion theorem gives us the following inversion formula:

Corollary 1 *If f, $\phi \in L$, then*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt.$$

Convoluting the distribution function F with the uniform distribution on [-h, h] and using Corollary 1 we can reproduce Levy's important theorem [15].

Corollary 2 *If f, $\phi \in L$, then*

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin ht}{ht} e^{-its} \phi(t) dt.$$

where L is the set of Lebesque integrable functions.

This result along with the work done by Gurland is then reworked by Davies [3] to produce a simple Numerical Inversion Formula for obtaining Cumulative Distribution Functions from Characteristic Functions. This result is reproduced and slightly reworked in Shepard's own paper [4] to obtain both Univariate and Multivariate Numerical Inversion Formulas. The final key results we reproduce here. In both Theorems $\delta(g, j)$ is defined so that $\delta(g, j)P(X \leq x) = P(X \leq x - (2\pi j/h_g)) - P(X > x + (2\pi j/h_g))$, where h_g should be thought of as the step size of the integration procedure.

Theorem 4.6.1. *(Univariate Case)* [6] For $h_1 > 0$

$$F(x) + \sum_{j=1}^{\infty} (-1)^j [\delta(1, j) P(X \leq x)]$$

$$= \frac{1}{2} - \frac{1}{\pi} h_1 \sum_{v=0}^{\infty} Im \left[\frac{\phi(h_1(v+\frac{1}{2})) \exp\left[-ixh_1(v+\frac{1}{2})\right]}{h_1(v+\frac{1}{2})} \right].$$

with the induced numerical integration error having the following absolute value,

$$\left| \sum_{j=1}^{\infty} (-1)^j [\delta(1,j) P(X \leq x)] \right| \leq max \left[F\left(x - \frac{2\pi}{h_1}\right), 1 - F\left(x + \frac{2\pi}{h_1}\right) \right].$$

The result is then extended to the Bivariate Case in a similar fashion.

Theorem 4.6.2. *(Bivariate Case)* [4] For $h_1, h_2 > 0$,

$$2^{-2} u(x_1, x_2) + \sum_{j=0}^{\infty} \left(\left[\prod_{k=1}^{2} (-1)^{jk} \delta(k, j_k) \right] F(x_1, x_2) \right)$$

$$= \frac{-2}{(2\pi)^2} h_1 h_2 \left[\sum_{v_1=0}^{\infty} \sum_{v_2=-\infty}^{\infty} Re \left[\frac{\phi(b_1, b_2) e^{-ix'b}}{b_1 b_2} \right] \right],$$

where $b = (h_1(v_1 + \frac{1}{2}), h_2(v_2 + \frac{1}{2}))' = (b_1, b_2)'$ *and* $u(x_1, x_2) = F(x_1, x_2) - 2[F(x_1) + F(x_2)] + 1$.

For our simulation we selected step size h = 0.001 for both distributions and summed over N = 50 iterations. This produced very good results in the univariate case, but due to the curse of dimensionality, resulted in poor estimates in the bivariate case. This was done due to computational constraints, and since we are only interested in a rough overview of the dependence structure, we did not attempt a more complex code that would speed up the procedure and allow smaller step sizes and larger iteration counts.

We also decided to set $\alpha = 0$ to simplify our work. This does make the bivariate simulation run somewhat faster, and after generating a few bivariate samples with $\alpha = 1$ without noticing any real difference in the correlation structure or shape of the distribution, other than the obvious shift caused by a change in location parameter, we decided this simplification was in order. Below we present graphs of the univariate and vivariate CDFs. Note since $\alpha = 0$, both our marginals now have the same CDF.

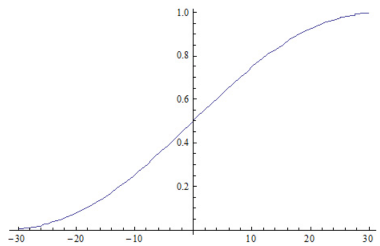

Figure 3: Univariate CDF of the DJ Marcus Marginals

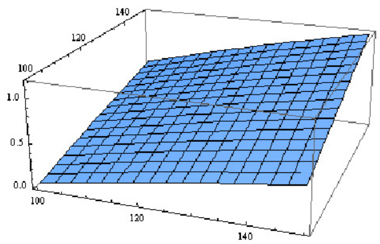

Figure 4: Bivariate CDF of the DJ Marcus Distribution I

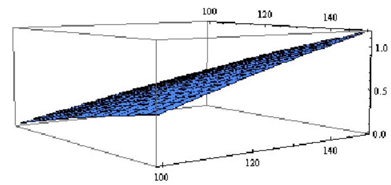

Figure 5: Bivariate CDF of the DJ Marcus Distribution II

Using these CDF estimates, we applied the discrete Inverse CDF method to generate 50 samples of size 50 each from the Bivariate DJ Marcus distribution. For each sample we calculated Pearson's Correlation, Schweizer and Wolff's Sigma and Szekely's Distance Correlation. We replicated the experiment to get a very good sample size for our analysis and in Figures 6 and 7 we plot each of these estimates.

As expected both Pearson's Rank Moment Correlation, and Schweizer and Wolff's Sigma do not converge properly, with estimates of the dependence measure ranging from .67 to .94 and .76 to 1.12. This shows that the lack of moments proves to be very troublesome for these two empirical estimates of dependence, with an inability to converge on any one particular value. Furthermore, the Empirical estimate of Scheweizer and Wolff's Sigma, are greater than 1 which should be the maximum attainable value. That may be due either to our poor random generation algorithm for the bivariate case, but also can be do to poor copula estimation for this type of distribution. Schweizer and Wolff's sigma does however seem to follow the Pearson's Rank Moment Correlation estimates very closely, a fact not noticed in Szekely's measure of dependence. This last measure is the most stable one that seems to converge around a value of .986, which is encouraging as it shows the measure has great potential for use in situations where only fractional moments less than 1 exist.

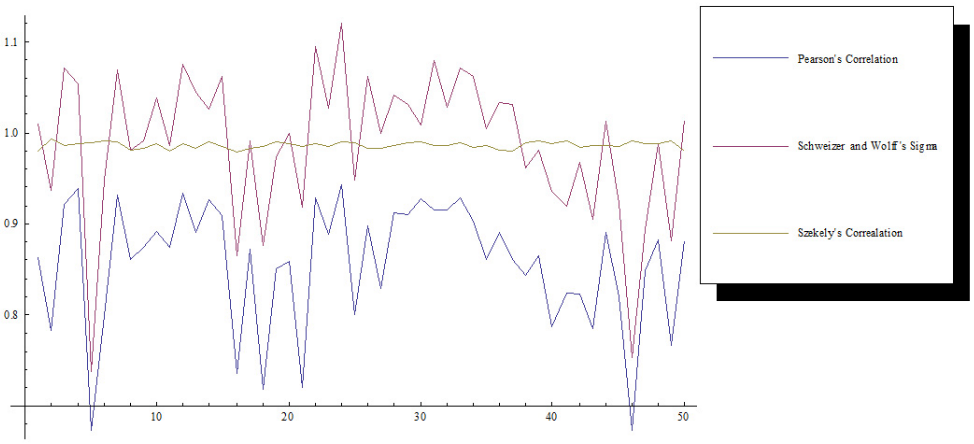

Figure 6: Three Dependence Measures of the DJ Marcus Distribution I

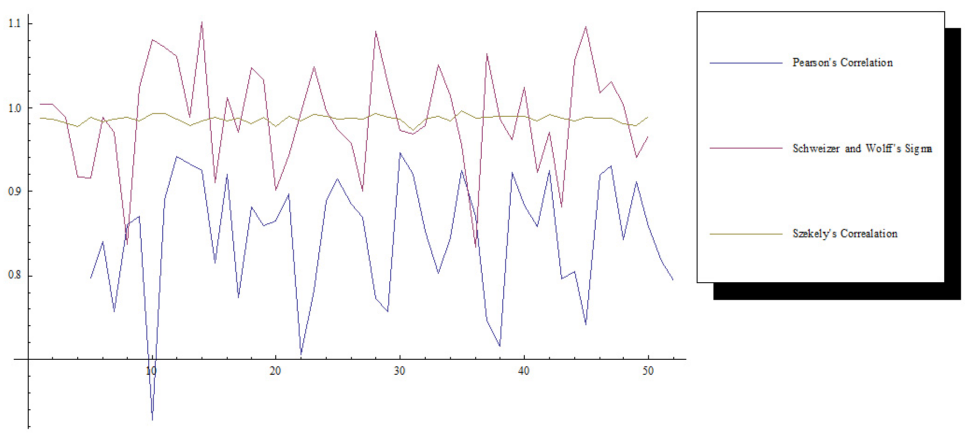

Figure 7: Three Dependence Measures of the DJ Marcus Distribution II

To further look into the behaviour of all three measures of dependence we decided to repeat the experiment, generating a further 50 sample of size 50 and taking a look at the difference between Pearson's Rank Moment Correlation and Schweizer and Wolff's Sigma, along the plot of the three measures individually.

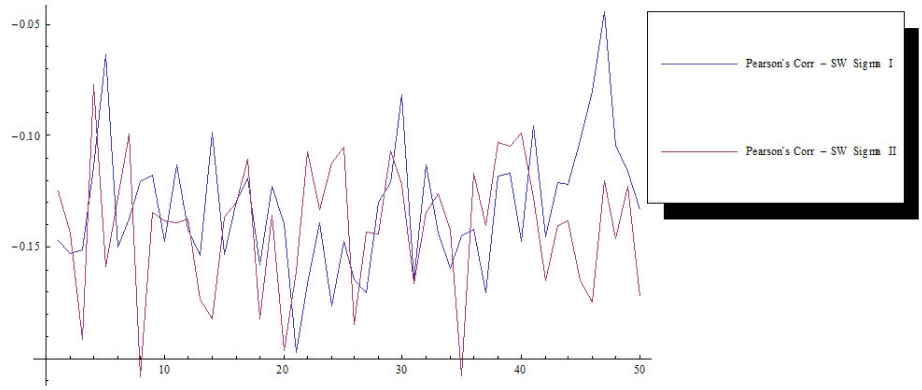

Figure 8: Difference between rho and SW Sigma for samples I and II

It is clear from these graphs that, as in the case of the bivariate normal distribution and the FGM bivariate exponential distribution, there may be an inequality we can define for this particular joint distribution function. It seems entirely possible that in this case $\rho^2 \leq R^2(X,Y) \leq \sigma^2_{X,Y}$, but this is of yet a conjecture.

5 A Dependence Measure for Lévy Processes

Cont and Tankov [2] introduced the concept of Lévy copulas in their book on Financial Modelling published in 2004. We reproduced the principal theorems and notions of this theory in the Mathematical Background section. Lévy copulas are an important tool for studying the dependence structure of Lévy processes, one of the few currently available. Traditional copulas fall short when directly applied, as highlighted by Cont and Tankov's work, and most other tools for analyzing dependence have yet to be adapted to Lévy processes. In this section, we will present a measure of dependence for Lévy processes based on Lévy Copulas following in the footsteps of Schweizer and Wolff, who introduced a dependence measure for random variables based on the classic copula concept. We will first cover positive Lévy copulas and then extend our theory to the general case since some modifications will be required.

5.1 Positive Lévy Processes

We start by deriving upper and lower bounds for positive Lévy copulas akin to the Frechet lower and upper bounds of classic copula theory. These will be used when testing the properties of our dependence measure, to ascertain if indeed it has all the properties of a measure of dependence.

Proposition 5.1.1. *Let $F(x,y) : [0,\infty]^2 \to [0,\infty]$ be a positive Lévy copula, therefore F is grounded and 2-increasing, as defined in Definition 2.2.2. Then for every pair (x, y) in* Dom F

$$0 \leq F(x,y) \leq \min(x,y).$$

Proof. $0 \leq F(x,y)$ by definition, and because positive Lévy copulas are 2-increasing and grounded they are also increasing in the marginals. Thus $F(x,y) \leq F(x,\infty) = x$ and $F(x,y) \leq F(\infty,y) = y$, so $F(x,y) \leq \min(x,y)$. □

It is worthwhile to note here that while at first glance the lower bound of zero may seem trivial this is actualy not the case. Cont and Tankov show that two Lévy processes are independent if and only if they never jump together, that is for every x and y such that $xy \neq 0$ the corresponding tail integral is zero. Hence, $F(x,y) = 0$ if x and y are finite. In other words, this trivial case actually represents the independence copula for Lévy processes with positive jumps, and in fact for general Lévy processes as well.

Furthemore, $\min(x,y)$ is also a Lévy copula, and in fact it is the dependence Lévy copula as shown by Cont and Tankov [2].

Definition 5.1.1. Let X_t and Y_t be two Lévy processes. Let the positive Lévy copula defined by their joint Lévy measure be $F : [0,\infty]^2 \to [0,\infty]$. Then the following is a measure of dependence of these two Lévy processes:

$$\sigma_{\nu_X,\nu_Y} = 2\int_0^\infty \int_0^\infty \frac{|F(x,y) - (x1_{y=\infty} + y1_{x=\infty})|}{e^x e^y} dxdy.$$

Here we choose to stick with Cont and Tankov's notation for the indepenence copula. Recall that Lévy copulas are functions of the joint Lévy measure, as such it essentially models how often the two processes of interest jump together. Therefore the maximum copula represents full dependence, while the zero copula, independence.

We will now show σ_{ν_X,ν_Y} has all the necessary properties to make it a measure of dependence between two Lévy processes. Since Cont and Tankov extended Sklar's theorem to Lévy copulas, we know that $F(x,y)$ exists and is uniquely defined on Ran $U_X\times$ Ran U_Y, the corresponding tail integrals; therefore, σ_{ν_X,ν_Y} is defined for every pair of Lévy processes X and Y. This also tells us that $\sigma_{\nu_X,\nu_Y} = \sigma_{\nu_Y,\nu_X}$, since $F(x,y)$ is uniquely defined on Ran $U_X \times$ Ran U_Y. Further, notice the above integral is zero if and only if $F(x,y) = x1_{y=\infty} + y1_{x=\infty}$ that is if and only if the unique copula defined for processes X_t and Y_t is in fact the independence copula. To check that $\sigma_{\nu_X,\nu_Y} = 1$, it suffices to use the upper bound on Lévy copulas $F(x, y) = \min(x, y)$ and plug it into our formula

$$
\begin{aligned}
\sigma_{\nu_X,\nu_Y} &= 2\int_0^{\infty}\int_0^{\infty} \frac{|\min(x,y) - (x1_{y=\infty} + y1_{x=\infty})|}{e^x e^y} dxdy \\
&= \lim_{T\to\infty} 2\int_0^{T}\int_0^{T} \frac{|\min(x,y) - (x1_{y=\infty} + y1_{x=\infty})|}{e^x e^y} dxdy \\
&= \lim_{T\to\infty} 2\int_0^{T}\int_0^{T} \frac{\min(x,y)}{e^x e^y} dxdy = 1.
\end{aligned}
$$

Thus the upper limit on the measure is 1 as needed. Now we must modify the dependence property of continuity when dealing with Lévy copulas. Namely since Lévy copulas are constructed using tail integrals, rather then distribution functions, the last condition of the dependence measures should read as follows:

- If (X_t, Y_t) and (X_{tn}, Y_{tn}) are pairs of Levy processes with joint Levy measures ν and ν_n, respectively, and if the sequences $\{\nu_n\}$ converges weakly to ν, then $\lim_{n\to\infty} \sigma_{\nu_{X_n},\nu_{Y_n}} = \sigma_{\nu_X,\nu_Y}$.

Proving this is a straighforward matter of noticing that since ν_n converges weakly to ν then by definition U_n converges weakly to U, the corresponding tail integrals. Since $U(x, y) = F(U_X(x), U_Y(y))$, then $F_n(x, y)$ converges weakly to $F(x, y)$. Since positive Levy copulas are bounded, then by the Lebesque dominated convergence theorem, we have

$$\lim_{n\to\infty} \sigma_{\nu_{X_n},\nu_{Y_n}} = \sigma_{\nu_X,\nu_Y}.$$

Thus, σ_{ν_X,ν_Y} has all the required properties to be defined as a measure of dependence, with the exception of invariance under monotone transformations. Since this invariance is not suited to Lévy processes, as infinite divisibility is destroyed under such transformations as shown by Cont and Tankov ([2] p.143), we do not seek this property for our given measure of dependence.

5.2 General Lévy Processes

We proceed in a similar manner as for positive Lévy copulas, proving all the necessary definitions required for our defined measure to be a measure of dependence.

Theorem 5.2.1. *Let $F(x,y) : [-\infty,\infty]^2 \to [-\infty,\infty]$ be a general Lévy copula. Then F is 2-increasing, $F(x,\infty) - F(x,-\infty) = F(\infty,x) - F(-\infty,x) = x$, and $F(0,x) = F(x,0) = 0\ \forall x$. Then for every pair (x, y) in the domain of F*

$$-\min(|x|,|y|)1_{xy\le 0} \le F(x,y) \le \min(|x|,|y|)1_{xy\ge 0}.$$

Proof. Notice that the trivial case when either $x = 0$ or $y = 0$ is obvious. Now, because F is 2-increasing, then for every rectangle $B = [x_1,x_2]\times[y_1,y_2] \subset [-\infty,\infty]^2$, the F volume of B, defined as $V_F(B) = F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1) \ge 0$. Pick some pair $(x,y) \in [-\infty,\infty]^2$ and define the following rectangles with the two points

$$B_1 = [0,x]\times[y,\infty], B_2 = [0,x]\times[-\infty,y],\ \text{if } x > 0,$$

$$B_3 = [x,0]\times[y,\infty], B_4 = [x,0]\times[-\infty,y],\ \text{if } x < 0.$$

Since $V_F(B) \ge 0$ for all the rectangles defined above, we have the following results:

$$V_F(B_1) \ge 0 \Rightarrow F(x,y) \le F(x,\infty) = x + F(x,-\infty).$$

But $F(x,-\infty) = F(U_X^+(z), U_Y^-(0))$, where $U_X^+(z) = x$, and $U_Y^-(0) = -\infty$. Furthermore, by definition $F(U_X^+(z), U_Y^-(0)) = U^{+-}(z,0) = -\nu([x,\infty)\times(-\infty,y])$. So for rectangle B_1 with $x > 0$, we have

$$F(x,y) \le x - \nu([x,\infty)\times(-\infty,y]) \le |x|.$$

Similarly, for rectangle B_2 with $x > 0$, we have

$$F(x,y) \geq F(x,-\infty) = -x + F(x,\infty) = -x + U^{++}(z,0)$$

$$= -x + \nu([x,\infty) \times [y,\infty)) \geq -|x|.$$

Finally, for rectangles B_3 and B_4 with $x < 0$, we have similar inequalities holding:

$$F(x,y) \geq F(x,\infty) = x + F(x,-\infty) = x + U^{--}(z,0)$$

$$= x + \nu((-\infty,x] \times (-\infty,y]) \geq -|x|,$$

$$F(x,y) \leq F(x,-\infty) = -x + F(x,\infty) = -x + U^{-+}(z,0)$$

$$= -x + \nu((-\infty,x] \times [y,\infty)) \leq |x|.$$

Thus, $\forall x \in [-\infty,\infty]$ we have $-|x| \leq F(x,y) \leq |x|$, and by a similar proof $\forall y \in [-\infty,\infty]$, we have $-|y| \leq F(x,y) \leq |y|$. Furthermore, notice that if $xy \leq 0$, then $F(x,y) \leq 0$ by definition, and if $xy \geq 0$, then $F(x,y) \geq 0$. This then gives us the upper and lower bounds on general Levy copulas and proves our theorem. $\square$

Notice that both bounds are general Lévy copulas corresponding to complete negative dependence for the lower bound, and complete positive dependence for the upper bound, as shown in Cont and Tankov [2].

Definition 5.2.1. Let X_t and Y_t be Lévy processes with joint Lévy measure ν. Let $F(x,y) : [-\infty,\infty]^2 \to [-\infty,\infty]$ be the general Lévy copula defined by this measure. Then the following is a measure of association for the two processes:

$$\sigma_{\nu_X,\nu_Y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{|F(x,y) - (x 1_{y=\infty} + y 1_{x=\infty})|}{e^{|x|}e^{|y|}} dxdy.$$

Proving that this is a measure of dependence follows a similar logic as in the case of positive Lévy copulas. Every pair of Lévy processes (X_t, Y_t) has an associated general Lévy copula defined for it, as shown by Cont and Tankov ([2] p.158). Thus, σ_{ν_X,ν_Y} is defined for every pair of Lévy processes. By the defintion of general Lévy copulas, the copula of (X_t, Y_t) is the same as the copula of (Y_t, X_t), and thus, σ_{ν_X,ν_Y} $= \sigma_{\nu_Y,\nu_X}$. Clearly, this measure is zero if and only if the general Lévy copula is the independence copula. Finally, if (X_t, Y_t) and (X_{tn}, Y_{tn}), $n = 1, 2, \ldots$ are pairs of Lévy processes, with joint Lévy measures ν and ν_n respectively, and if the sequence $\{\nu_n\}$ converges weakly to ν, then $\lim_{n\to\infty} \nu_n = \nu$. This follows from the definition of general Lévy copulas, which are functions of the respective joint Lévy measures. Thus, if $\{\nu_n\}$ converges weakly to ν, then $F_n(x, y)$ converges weakly to $F(x, y)$. Since general Lévy copulas are bounded, then by the Lebesgue dominated convergence theorem, we have the needed assertion.

6 Conclusions and Future Work

In this thesis we presetend two classic dependence measures for random variables based on copulas and characteristic functions, respectively. The reader was introduced to Renyi's updated properties of dependence measures, and based on these, we extended both measures to the multivariate case. We compared them using specific bivariate distributions and established inequalities between these measures and the classic Pearson's product moment correlation for each family of bivariate distributions. In particular we applied these dependence measures to random variables without first moments, studying their behaviours under conditions for which convergence has not yet been proven, learning that at least one measure, based on characteristic functions, can indeed be modified to be used in such cases.

We took our work for analyzing the association between random events to the case of stochastic random variables. In particular we introduced the reader to the concept of Lévy processes, and of Lévy copulas, which are becoming popular in studying the dependence structure of such processes. We then introduced a brand new measure of dependence for both positive and general Lévy processes, applying all of Rényi's properties where such properties made sense.

6.1 Mutual Information Index

In the future we would like to revisit our classic measures of dependence for random variables, by examining at a new measure gaining popularity and based on Shannon's Information Theory. This measure of dependence looks at the difference between the multivariate density function and the product of its marginals, in essence, measuring the difference between the entropy contained by a given multivariate distribution and that contained by a fully independent set of random variables. Shannon's Entropy has found a wide range of applications across the breadth of scientific fields. Mutual

Information in particular has stirred interest as a replacement to Pearson's Rank Moment Correlation primarily because of its ability to capture non-linear depenedence structures, which can often be found especially in financial models. Dionisio, Menezes and Mendes [5] have a very good introduction of this in their paper. We define the Entropy of a given random variable X as well as that of a bivariate distribution as

$$
\begin{aligned}
H_X \;\; &= -E[\log(p_X(x))] = -\int p_X(x)\log(p_X(x))dx, \\
H_{XY} &= -E[\log(p_{XY}(xy))] = -\int\int p_{XY}(xy)\log(p_{XY}(xy))dxdy.
\end{aligned}
$$

Then we can define Mutual Information in the following way

$$I_{XY} = H_X + H_Y - H_{XY}.$$

Most authors use this definition of Mutual Information using it as a measure of correlation or "information" in a given system. This is, however, not satisfactory for our purposes since it can easily be seen to be greater then one, which then means it does not fit the outline properties we seek for a good measure of dependence. Fortunately, this is easily remedied by defining our Mutual Information Index as follows

$$\lambda_{XY} = \sqrt{1 - e^{-2I_{XY}}}.$$

Naturally, this can quite easily be extended to the multivariate case as follows

$$I_{X_1 \ldots X_k} = H_{X_1} + \ldots + H_{X_k} - H_{X_1 \ldots X_k},$$

$$\lambda_{X_1 \ldots X_k} = \sqrt{1 - e^{-2I_{X_1 \ldots X_k}}},$$

where

$$
\begin{aligned}
H_{X_1} &= -\int_{R^d} p_{X_1}(x_1) \log\left(p_{X_1}(x_1)\right) dx_1, \\
H_{X_1 \dots X_k} &= -\int \dots \int p_{X_1 \dots X_k}(x_1 \dots x_k) \log\left(p_{X_1 \dots X_k}(x_1 \dots x_k)\right) dx_1 \dots dx_k.
\end{aligned}
$$

The measure looks promising as a potential path for analyzing multipoint correlations, and giving a meaningful explanation to what they entail, however there are certain issues which make applying this measure difficult. We often do not have explicit density functions available, especially in the case of more complicated stochastics, and the convergence of estimates can be quite slow and unwieldy to be very practical. We would like to design and implement an algorithm for applying this method to the more volatile random events, primarily found in finance and atmospheric modelling.

6.2 Lévy Dependence Application

The dependence measure we introduced for Lévy processes has very nice theoretical properties, and seems to fit most of the Rényi conditions for being a dependence measure. We would like, in the future, to apply this measure to an actual data set and see how it performs. In particular, we are very interested in an existing problem within wind modelling. To build good predictive models for wind speeds in a particular area, one often needs a good deal of data gathered over an extensive period of time. A minimum of one year of wind data is required for proper modelling within a given area [1] but often at least ten years are needed for truly accurate predictions of wind potential over the usual twenty to thirty year lifetime of a wind farm [19].

Often times such data is not available at the site of interest, however, one can find decades of meteorological data at nearby sites, such as weather stations, radio stations, or airports. A method for measuring the dependence between the two sites, modelling the joint behaviour of wind across the two regions, and using past data from one site to build a more thorough wind model at the site of interest should be developed. We propose fitting Lévy processes to wind data available at both sites, and joining these two with Lévy copulas. We can then use our measure of dependence to see if there exists a strong correlation between the two sites of interest.

If such a correlation does exist, one can then use the past data at site two to build a more effective wind model at the first site of interest getting better prediction for total energy potential, and thus, the economic viability of the site for wind energy production.

Appendices

A Other Metrics for Probability Distributions

Measures of Correlation and Dependence are one way to look at the relationship between random variables. In essence every such measures looks at how two random events vary together. Constructing measures of dependence between random variables X and Y given the random vector (X, Y) requires using metrics between (X, Y) and (X', Y'), where $X \sim X'$ and $Y \sim Y'$ in distribution, and X', Y' are independent. In this dissertation we have used a few such metrics, e.g. absolute value of the difference between the characteristic functions of (X, Y) and (X', Y'), but many other options exist.

In our future work, we would like to try and produce other measures of dependence based on different metrics, and thus, careful consideration needs to be give to the type of metric used. While this is not the focus of our thesis, we provide a summary and a few new results regarding bounds on various possible choices of metrics. Knowledge of different metrics can often provide bounds for the ones we choose, and applying different metrics to the same problem can often give new insights into the random behaviour we are studying.

The purpose of this section is not only to present several such metrics, but to try and present relationships between them, something that is not often found in literature. The work is not our own, but was undertaken by Gibbs and Su [10] in an elegant and well written paper on the sujbect, and we summarize the results below. Let Ω denote a measureable space with σ-algebra B. Let M be the space of all probability measures on (Ω, B). Table 1 shows a summary of these ten metrics.

Abbreviation	Metric
D	Discrepancy
H	Hellinger Distance
I	RElative entropy (or Kullback-Leibler divergence)
K	Kolmogorov (or Uniform) metric
L	Levy metric
P	Prokhorov metric
S	Separation distance
TV	Total variation distance
W	Wasserstein (or Kantorovich) metric
X^2	X^2 distance

Table 1: Abbreviations for metrics used in Figure 1

Let μ, ν denote probability measures on Ω. Let f and g denote their corresponding density functions with respect to σ-finite dominating measure λ (for example, λ can be taken to be $(\mu+\nu)/2$). If $\Omega = \mathbf{R}$, let F, G denote their corresponding distribution functions. X and Y will denote random variables on Ω such that $L(X) = \mu$ and $L(Y) = \nu$. If Ω is a metric space, it will be measurable with the Borel σ-algebra. If Ω is a bounded metric space with metric d, then $diam(\Omega) = \sup\{d(x,y) : x, y \in \Omega\}$ denotes the *diameter* of Ω.

Kolmogorov (or Uniform) metric.

1. State space: $\Omega = \mathbf{R}$.
2. Definition:

$$d_K(F,G) := \sup_x |F(x) - G(x)|, x \in \mathbf{R}.$$

3. It assumes values in [0, 1] and is invariant under all increasing one-to-one transformations of the line.

Relative entropy (or Kullback-Leibler divergence)

1. State space: Ω any measurable space.

2. Definition: if f, g are densities of the measures μ, ν with respect to a dominating measure λ,

$$d_I(\mu, \nu) := \int_{S(\mu)} f \log{(f/g)} d\lambda,$$

where $S(\mu)$ is the support of μ on Ω. For a countable space Ω,

$$d_I(\mu, \nu) := \sum_{\omega \in \Omega} \mu(\omega) \log \frac{\mu(\omega)}{\nu(\omega)}.$$

3. Relative Entropy assumes values on $[0, \infty]$ and is not a metric; however, if $\mu = \mu_1 \times \mu_2, \nu = \nu_1 \times \nu_2$ on a product space $\Omega_1 \times \Omega_2$,

$$d_I(\mu, \nu) = d_I(\mu_1, \nu_1) + d_I(\mu_2, \nu_2).$$

Hellinger distance

1. State space: Ω any measurable space.

2. Definition: if f, g are densities of the measures μ, ν with respect to any dominating measure λ,

$$d_H(\mu, \nu) := \left[\int_\Omega (\sqrt{f} - \sqrt{g})^2 d\lambda\right]^{1/2} = \left[2\left(1 - \int_\Omega \sqrt{fg}\, d\lambda\right)\right]^{1/2}.$$

For a countable state space Ω,

$$d_H(\mu, \nu) := \left[\sum_{\omega \in \Omega} \left(\sqrt{\mu(\omega)} - \sqrt{\nu(\omega)}\right)^2\right]^{1/2}.$$

3. Assumes values on $[0, \sqrt{2}]$ and for product measures $\mu = \mu_1 \times \mu_2, \nu = \nu_1 \times \nu_2$ on a product space $\Omega_1 \times \Omega_2$,

$$1 - \frac{1}{2} d_H^2(\mu, \nu) = \left(1 - \frac{1}{2} d_H^2(\mu_1, \nu_1)\right)\left(1 - \frac{1}{2} d_H^2(\mu_2, \nu_2)\right).$$

Lévy metric.

1. State space: $\Omega = \mathbf{R}$.

2. Definition:

$$d_L(F, G) := \inf\{\epsilon > 0 : G(x - \epsilon) - \epsilon \leq F(x) \leq G(x + \epsilon) + \epsilon, \forall x \in R\}.$$

3. It takes values on [0, 1], is shift invariant, but not scale invariant.

Prokhorov (or Lévy-Prokhorov) metric.

1. State space: Ω any metric space.

2. Definition:

$$d_P(\mu, \nu) := \inf\{\epsilon > 0 : \mu(B) \leq \nu(B^\epsilon) + \epsilon \text{ for all Borel sets } B\},$$

   where $B^\epsilon = \{x : \inf_{y \in B} d(x, y) \leq \epsilon\}$. It assumes values in [0, 1].

Separation distance.

1. State space: Ω a countable space.

2. Definition:

$$d_S(\mu,\nu) := \max_i \left(1 - \frac{\mu(i)}{\nu(i)}\right).$$

3. This is not a metric, and assumes values in [0, 1].

Total variation distance.

1. State space: Ω any measurable space.

2. Definition:

$$d_{TV}(\mu,\nu) := \sup_{A\subset\Omega} |\mu(A) - \nu(A)| = \frac{1}{2}\max_{|h|\leq 1}\left|\int h d\mu - \int h d\nu\right|,$$

where $h : \Omega \to R$ satisfies $|h(x)| \leq 1$. This metric assumes values in [0, 1] For a countable state space Ω,

$$d_{TV} := \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|,$$

which is half the L^1-norm between the two measures.

Wasserstein (or Kantorovich) metric.

1. State space: $\mathbf{R}$ or any metric space.

2. Definition: For $\Omega = \mathbf{R}$, if F, G are the distribution functions of μ, ν respectively, the *Kantorovich metric* is defined by

$$d_W(\mu,\nu) := \int_{-\infty}^{\infty} |F(x) - G(x)|dx = \int_0^1 |F_1(t) - G^{-1}(t)|dt,$$

where F^{-1}, G^{-1} are the inverse functions of F, G. For any separable metric spaces:

$$d_W(\mu,\nu) := \sum\left\{\left|\int h d\mu - \int h d\nu\right| : \|h\|_L \leq 1\right\},$$

the supremum being taken over all h satisfying the Lipschitz condition $|h(x) - h(y)| \leq d(x,y)$, where d is the metric on Ω.

3. This metric takes values in $[0, \operatorname{diam}(\Omega)]$, where $\operatorname{diam}(\Omega)$ is the diameter of the metric space (Ω, d). Furthermore, by the Kantorovich-Rubinstein theorem, the Kantorovich metric is equal to the *Wasserstein metric*

$$d_W(\mu,\nu) = \inf_J\{E[d(X,Y)] : L(X) = \mu, L(Y) = \nu\},$$

where the infimum is taken over all joint distributions J with marginals μ, ν.

X^2-distance.

1. State space: Ω any measurable space.

2. Definition: if f, g are densities of the measures μ, ν with respect to any dominationg measure λ, and $S(\mu), S(\nu)$ are their supports on Ω,

$$d_{X^2}(\mu,\nu) := \int_{S(\mu)\cup S(\nu)} \frac{(f-g)^2}{g} d\lambda.$$

For countable Ω this reduces to:

$$d_{X^2}(\mu,\nu) := \sum_{\omega\in S(\mu)\cup S(\nu)} \frac{(\mu(\omega)-\nu(\omega))^2}{\nu(\omega)}.$$

3. The X^2-distance is not a metric and takes values in $[0, \infty]$.

Gibbs and Su give a detailed accounting of the relationship between these metrics, summarizing those found throughout the literature on the subject, as well as proving theorems introducing new bounds on some of these metrics. These results are summarized in a diagram showing the relationship between all ten metrics, which we reproduce below in Figure 1.

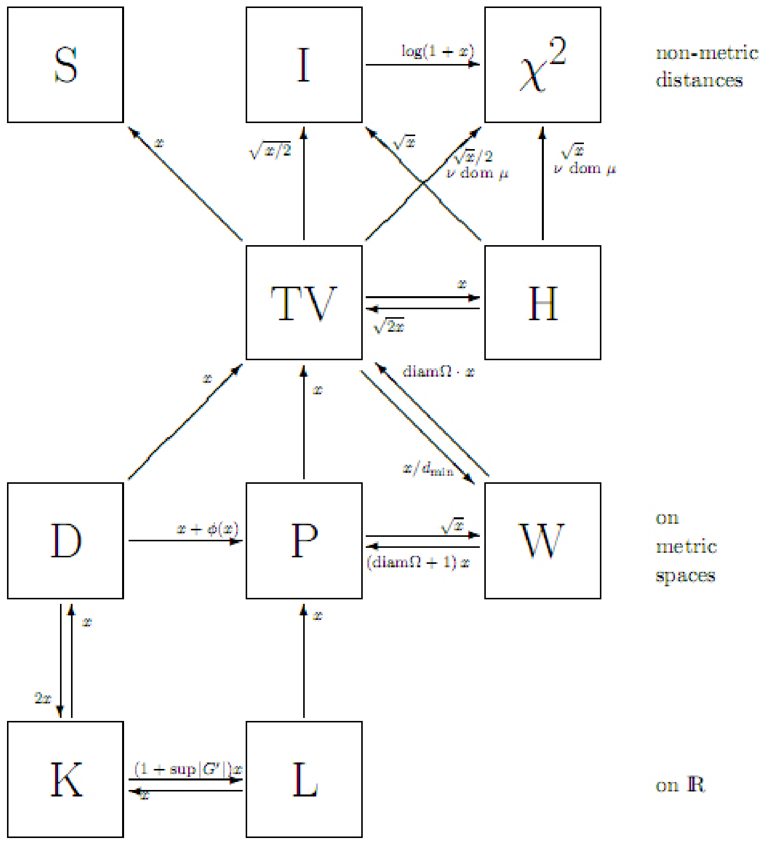

Figure 1. Relationships among probability metrics. A directed arrow from A to B annotated by a function $h(x)$ means that $d_A \leq h(d_B)$. The symbol diam Ω denotes the diameter of the probability space Ω; bounds involving it are only useful if Ω is bounded. For Ω finite, $d_{min} = \inf_{x,y \in \Omega} d(x, y)$ (from Gibbs and Su [10]).

Appendices

B Overview of Algorithms and Code

In order to test the measures of dependence presented in this thesis, we developed and implemented several algorithms in R and Mathematica. Below we outline these algorithms along with a short explanation for what task they perform.

B.1 Inversion Algorithm

The first algorithm we present, is a random number generator, based on the inversion of characteristic functions. A pseudo-code for the algorithm can be found in [26] and it serves as the basis of our Mathematica code. The algorithm is used to generate 50 samples from a bivariate non-stable distribution with stable marginals, better known in our thesis as the D.J. Marcus example.

```
##Code for generating 50 samples from a bivariate distribution with
##stable marginals.  a is between 0.5 and 1 not inclusive


h=0.0001 ##the "step size" of the algorithm, smaller the better
         ##the approximation, and longer the runtime
a=0.51 ##the "alpha" parameter of the stable marginal distributions
T = 50 ##Constant used to determine the range over which our
      ##distribution function is to be estimated in the bivariate case


##THe code below estimates univariate and bivariate CDFs for the
##D.J. Marcus distribution
##mu = 0.  We set the mean of our bivariate distribution to zero,
##for ease without any loss of generality
```

```
##First Marginal Estimation

F[x_]:=0.5 - (1/Pi)*h*Sum[ Exp[-h^a * (v + 0.5)^a]*

           *Sin[-x*h*(v+0.5)]/(h*(v+0.5)),{v, 0, 1000}]

##Second Marginal Estimation

G[y_]:=0.5 - (1/Pi)*h*Sum[ Exp[-h^a * (v + 0.5)^a]*

           *Sin[-y*h*(v+0.5)]/(h*(v+0.5)),{v, 0, 1000}]


##The three functions below combined to give an estimate for the

##bivariate CDF.  L[x,y] represents the kernel of the inversion

##formula K[x,y] sums over the Kernel to further refine the estimate,

##and H[x,y] finally calculates the actual bivariate CDF.


L[x_,y_]:=Exp[-(h^2*(i+0.5)^2 + h^2*(j+0.5)^2)^(a/2)]*

                     *Cos[h*(i + 0.5)*x + h*(j + 0.5)*y]

                                   / (h*(i+0.5)*h*(j+0.5))

K[x_,y_]:=-2*h^2 * Sum[ Sum[ L[x,y], {j, -T, T}], {i, 0, T}]

H[x_,y_]:= K[x,y] - 0.5*(F[x] + G[y]) + 0.25



##The code below calculates the CDF of X and the conditional CDF of Y

##based on the above defined functions.  It is finally used to

##generate the random sample of interest from the bivariate

##D.J. Marcus distribution

CDFX = Table[F[i], {i, -30, 30}]


##X takes values between -30 and 30

x = RandomInteger[{1, 61}]
```

```
X = {x - 31}


##Y must take values between 100 and 160, due to the properties of
##the bivariate D.J. Marcus distribution
CondCDFY = CDFX[[x]]*H[x+98, RandomInteger[{99, 160}]]


n = -30; While[CondCDFY > F[n] && n < 30, (y = n+1; n++)]
Y = {y}


For[i=1, i<50, i++, x = RandomInteger[{1, 61}];
        X = Append[X, x-31];
        CondCDFY = CDFX[[x]]*H[x+98, RandomInteger[{99, 160}]];
        n = -30;
        While[CondCDFY > F[n] && n < 30, (y = n+1; n++)];
               Y = Append[Y, y] ]



##mu = 1.  We implement the same functions above for the case when
##the mean of the bivariate D.J. Marcus distribution is 1,
##noticing no real change in the correlation structure as expected.
F[x_]:=0.5 - (1/Pi)*h*Sum[ Exp[-h^a * (v + 0.5)^a]
           *Sin[-x*h*(v+0.5)]/(h*(v+0.5)),{v, 0, 1000}]
G[y_]:=0.5 - (1/Pi)*h*Sum[ Exp[-h^a * (v + 0.5)^a]
           *Sin[(1 - y)*h*(v+0.5)]/(h*(v+0.5)),{v, 0, 1000}]
L[x_, y_]:= Exp[-(h^2*(i+0.5)^2 + h^2*(j+0.5)^2)^(a/2)]*
           *Cos[ Sqrt[h^2*(i+0.5)^2 + h^2*(j+0.5)^2]
           *Cos[ArcTan[(j+0.5)/(i+0.5)]] - (x*h*(i+0.5)
```

```
            + y*h*(j+0.5))] / (h^2*(i+0.5)*(j+0.5))
K[x_,y_]:=-2*h^2 * Sum[ Sum[ L[x,y], {j, -T, T}], {i, 0, T}]
H[x_,y_]:= K[x,y] - 0.5*(F[x] + G[y]) + 0.25


##Plotting the Correlations
Needs["PlotLegends`"]
ListLinePlot[{rho2, sigma2, drho2},
        PlotLegend -> {"Pearson's Correlation",
             "Schweizer and Wolff's Sigma", "Szekely's Correalation"}]
ListLinePlot[{rho, sigma, drho},
        PlotLegend -> {"Pearson's Correlation",
             "Schweizer and Wolff's Sigma", "Szekely's Correalation"}]
ListLinePlot[{diff1, diff2},
        PlotLegend -> {"Pearson's Corr - SW Sigma I",
                          "Pearson's Corr - SW Sigma II"}]
```

B.2 Empirical Dependence Algorithms

Once we had generated our random sample, we set about calculating Pearson's Correlation, Schweizer and Wolff's Sigma and Szekely's Dependence for the entire sample. The data was entered directly into R separately, and stored in a data frame. Notice that we had fifty samples of size fifty each, and thus were able to calculate fifty different measures of dependence.

It is worthwhile to note that the Schweizer and Wolff's sigma gave us the most trouble. The original algorithm we designed, used a "for" loop at it's core, which turned out to be extremely inefficient and slow in computation time. A second algorithms had to be designed, splitting up the for loop into one big vector of data over which to estimate our S.W. sigma. We included both versions, because the second version, while orders of magnitude faster, is far more difficult to understand then the original.

B.2.1 Data Setup and Pearson's Correlation

```
##Setting up our data set for processing and calculating Pearson’s
##product moment.
data = data.frame(X1, Y1, X2, Y2, X3, Y3, X4, Y4, X5, Y5, X6, Y6,
           X7, Y7, X8, Y8, X9, Y9, X10, Y10, X11, Y11, X12, Y12,
           X13, Y13, X14, Y14, X15, Y15, X16, Y16, X17, Y17, X18, Y18,
           X19, Y19, X20, Y20, X21, Y21, X22, Y22, X23, Y23, X24, Y24,
           X25, Y25, X26, Y26, X27, Y27, X28, Y28, X29, Y29, X30, Y30,
           X31, Y31, X32, Y32, X33, Y33, X34, Y34, X35, Y35, X36, Y36,
           X37, Y37, X38, Y38, X39, Y39, X40, Y40, X41, Y41, X42, Y42,
           X43, Y43, X44, Y44, X45, Y45, X46, Y46, X47, Y47, X48, Y48,
           X49, Y49, X50, Y50)

##Estimates of the Correlation
rho = cor(data[,1], data[,2])
for (i in seq(3, 100, 2))
rho = c(rho, cor(data[,i], data[,i+1]))
```

B.2.2 First Schweizer and Wolff's Sigma Algorithm

```
##The original version of the Schweizer and Wolff’s Sigma.
##Function C calculates C(i/n, j/n) = number of pairs (x,y) in the
##sample with x <= ith order statistic of x, and y <= jth order
##statistic of y
C <- function(x,y) {
```

```
        temp = expand.grid(sort(x), sort(y))
        a = rep(temp[,1], each=length(x))
        b = rep(temp[,2], each=length(y))
        temp2 = identity(x <= a) * identity(y <= b)
        count = sum(temp2[1:50])
        for(i in c(2:2500)) {
            l = i*50 - 49
            k = i*50
            count = c(count, sum(temp2[l:k]))
        }
        count/length(x)
}

##Function F calculates the double sum of i/n times j/n for i and j
##going from 1 to n.
F <- function(x,y) {
        a = rep(1:length(x))
        b = rep(1:length(y))
        apply(expand.grid(a,b),1,prod) / length(x)^2
}

##swSigma function combines C and F to calculate the SW sigma for a
##given sample (X,Y)
swSigma <- function(x, y)
        sum(abs(C(x, y) - F(x, y))) * 12 / (length(x)^2 - 1)

##Calculating the SW sigma for all 50 pairs of samples
```

```
sigma = swSigma(data[,1], data[,2])
for (i in seq(3, 100, 2))
       sigma = c(sigma, swSigma(data[,i], data[,i+1]))
```

B.2.3 Second Schweizer and Wolff's Sigma Algorithm

```
##The second SW sigma algorithm, collapsing the for loop into vector
##operations.  Functions C and F calculate the same things.
C <- function(x,y) {
       temp = expand.grid(sort(x), sort(y))
       a = rep(temp[,1], each=length(x))
       b = rep(temp[,2], each=length(y))
       sum(identity(x <= a) * identity(y <= b)) / length(x)
}

F <- function(x,y) {
       a = rep(1:length(x))
       b = rep(1:length(y))
       sum(apply(expand.grid(a,b),1,prod) / length(x)^2)
}

swSigma <- function(x, y)
       (C(x, y) - F(x, y)) * 12 / (length(x)^2 - 1)

sigma = swSigma(data[,1], data[,2])
for (i in seq(3, 100, 2))
       sigma = c(sigma, swSigma(data[,i], data[,i+1]))
```

B.2.4 Szekely's Dependence Algorithm

This code follows the algorithm outlined in Szekely's et al. paper [30].

```
##The lmean, rmean, dmean, and tmean correspond to ak., a.l, a.., and
##Akl respectively in Zzekely's paper.  The bk., b.l, b.., and Bkl
##functions are exactly the same and thus are not reiterated


akl <- function(x)
         abs(expand.grid(x,x)[,1] - expand.grid(x,x)[,2])



ak <- function(x) {
         temp = sum(abs(x[1] - x))
         for (i in c(2:length(x)))
                 temp = c(temp, sum(abs(x[i] - x)))
         temp / length(x)
         rep(temp, length(x))
}


al <- function(x){
        temp = sum(abs(x - x[1]))
        for (i in c(2:length(x)))
                 temp = c(temp, sum(abs(x - x[i])))
        temp / length(x)
        rep(temp, each=length(x))
}


a <- function(x)
```

```
        sum(abs(expand.grid(x,x)[,1] - expand.grid(x,x)[,2]))
                                                / length(x)^2


AKL <- function(x)
        akl(x) - ak(x) - al(x) + a(x)

dCov <- function(x, y)
        sum(AKL(x)*AKL(y)) / (length(x)*length(y))

dVar <- function(x)
        sum(AKL(x)^2) / length(x)^2

dCorr <- function(x,y)
        if (dVar(x)*dVar(y) == 0) 0 else dCov(x,y)
                                  / sqrt(dVar(x)*dVar(y))

drho = dCorr(data[,1], data[,2])
for (i in seq(3, 100, 2))
        drho = c(drho, dCorr(data[,i], data[,i+1]))
```

References

[1] S. Clarke. Electricity generation using small wind turbines at your home or farm. Fact Sheet 767, Ontario Ministry of Agriculture, Food and Rural Affairs, September 2003.

[2] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Financial Mathematics Series. Chapman & Hall/CRC, 2004.

[3] R. B. Davies. Numerical inversion of a characteristic function. *Biometrika*, 60:415–417, 1973.

[4] V. H. de la Pena, R. Ibragimov, and S. Sharakhmetov. Characterization of joint distributions, copulas, information, dependence, and decoupling, with applications to time series. *IMS Lecture Notes, 2nd Lehmann Symposium - Optimality*, 49:183–209, 2006.

[5] A. Dionisio, R. Menezes, and D. A. Mendes. Mutual information: a measure of dependency for nonlinear time series. *Physica A: Statistical Mechanics and its Applications*, 344(1-2):326–329, 2004.

[6] K. Dmowski. A multipoint correlation method with binomial weighting coefficients for deep-level measurements in metal-oxide-semiconductor devices. *Journal of Applied Physics*, 71, 1992.

[7] P. Embrechts, R. Frey, and A. J. McNeil. *Quantitative Risk Management: Concepts, Techniques, Tools*. Princenton University Press, 2005.

[8] E. W. Frees and E. A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 1998.

[9] C. Genest and L. Rivest. Statistical inference procedures for archimedean copulas. *Journal of the American Statistical Association*, 88:1034 – 1043, 1993.

[10] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

[11] J. Gil-Pelaez. Note o the inversion theorem. *Biometrika*, 37:481–482, 1951.

[12] J. Gurland.

[13] H. Joe and J. J. Xu. The estimation method of inference functions for margins for multivariate models. Technical report, Department of Statistics, University of British Columbia, 1996.

[14] J. Komelji and M. Perman. Joint characteristic functions constrution via copulas. *Insurance: Mathematics and Economics*, 47:137 – 143, 2010.

[15] P. Levy. *Calcul des probabilites*. 1925.

[16] D. J. Marcus. Non-stable laws with all projections stable. *Z. Wahrscheinlickhtkeitstheorie verw.*, pages 139–156, 1983.

[17] R. J. Miller, L. P. Dasi, and D. R. Webster. Multipoint correlations of concentration fluctuations in a turblent passive scalar field. *Experiments in Fluids*, 44, 2008.

[18] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer Science and Business Media, LLC, 2006.

[19] H. Nfaoui, J. Buret, and A. A. M. Sayigh. Wind characteristics and wind energy potnetial in morocco. *Solar Energy*, 63:51–60, 1998.

[20] M. Niemann and H. Kantz. Joint probability distributions and mutlipoint correlations of the continuous-time random walk. *Physical Review E*.

[21] A. P. Prudnikov, A. Brychkov, and O.I. Marichev. *Integrals and Series*. New York: Gordon and Breach Science Publishers, 1986.

[22] A. Renyi. On measures of dependence. *Acta Mathematica Hungarica*, 10:441–451, 1959.

[23] S. M. Ross. *Stochastic Processes*. Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., 1996.

[24] K. I. Sato. *Levy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.

[25] B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879 – 885, 1981.

[26] N. G. Shepard. From characteristic function to distribution function: a simple framework. *Econometric Theory*, 7:519–529, 1991.

[27] N. G. Shepard. Numerical integration rules for multivariate inversions. *J. of Statist. Comput. Simul.*, 39:37–46, 1991.

[28] A. Sklar. Fonctions de répartition á n dimensions et leurs marges. *Publ Inst Statist Univ Paris*, 8:229 – 231, 1959.

[29] A. Sklar. Sur une mesure de dépendence pour les variables aléatoires. *Publications de l'Institute de Statistique de l'Université de Paris*, 8:229–231, 1959.

[30] G. J. Szekely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

[31] N. Veraverbeke. Multivariate data and copulas. Lecture notes, Haselt Univeristy, 2005.

[32] F Šanda and S Mukamel. Multipoint correlation functions for continuous-time random walk models of anomalous diffusion. *Physical Review E*, 72:031108, 2005.

[33] E. F. Wolff. N-dimensional measures of dependence. *Stochastica*, 4:175–188, 1981.