An Information Theoretic Approach to Copulas

Ashler Herrick
Candidate for the B.S. Degree
in Applied Mathematics and Economics

State University of New York, College at Oswego
College Honors Program

May, 2020

Abstract

Information Theory for the most part is based on these central concepts; Entropy, Conditional Entropy, and Mutual Information. In the case of continuous random variables, the density function is used in the evaluation of these central concepts. Likewise, in multivariate situations, the joint density function is used in the evaluation. But, in most of the practical situations the joint density function is not that easy to formulate or even estimate. The situation may get worse when the sample size is small and the dimension is high. In such situations, one can use the Copulas to model the joint density function. The Copulas require only the marginal density functions and these are easier to formulate (or estimate). Our interest is in checking to see whether this Copula based approach is reasonable and if so, can Information Theory provide a new avenue to compare and analyze copula models.

Contents

1 Advice to Future Honors Students

The single piece of advice I can is to choose a topic that you enjoy. I thoroughly enjoyed the process of learning, exploring and writing about my topic. Without this interest, the quality of my thesis would have suffered greatly.

2 Acknowledgements

I would like to thoroughly acknowledge Dr. Nanthakumar for his efforts, not only on this paper, but throughout the many classes I have had the pleasure of being a student of his. Dr. Nanthakumar's enthusiasm for probability and statistics was so great that it sparked my own interest in the subject. He has provided me with guidance, advice and inspiration since I took my first mathematical statistics class with him. It is truly an honor to have him as an advisor to my thesis.

3 Authors Reflections

I am thankful that I had the opportunity to work with my advisors Dr. Nanthakumar and Dr. Jnawali on a paper that I am proud of, and a paper that allowed me to realize that I enjoy the research process. This realization solidified my desire to attend graduate school for applied mathematics.

4 Thesis Body

4.1 Preliminary Information Theory Definitions

The central concepts in Information theory are: Entropy, Conditional Entropy and Mutual Information. These formulas were originally defined only for discrete random variables but the concepts subsequently been extended to continuous random variables as well, this is typically referred as differential entropy.

Entropy can be thought of as the amount of uncertainty of a random variable. For a continuous random variable X it is defined as

$$H(X) = -\int_S f(x)\ln(x)dx$$

where S is the support of X and $f(x)$ is the p.d.f. If X is a continuous random variable the entropy is often referred to as differential entropy. In the discrete case, entropy is strictly positive, but this is not necessarily the case with differential entropy. An interesting note is that by the definition of expected value, we can write the previous definition in either case as

$$H(X) = E\left[\log\left(\frac{1}{f(x)}\right)\right],$$

where $f(x)$ is the p.m.f in the discrete case and the p.d.f in the continuous case.

Joint Entropy is similarly defined as

$$H(X,Y) = -\int_{S_y}\int_{S_x} f(x,y)\ln(f(x,y))dxdy = E\left[\ln\left(\frac{1}{f(x,y)}\right)\right]$$

where S_x, S_y are the support of X, Y respectively and $f(x,y)$ is the joint p.d.f. Similarly to univariate entropy, this quantity can be negative in the context of continuous random variables.

Kullback-Leibler Divergence, also referred to as relative entropy, is defined as

$$D(f||g) = \int_S f\log\frac{f}{g},$$

where f and g are probability density functions.

The mutual information of X, Y is defined as

$$I(X;Y) = \int_{S_y} \int_{S_x} f(x,y) \ln \left(\frac{f(x,y)}{f(x)f(y)} \right) dx dy = E \left[\ln \left(\frac{f(x,y)}{f(x)f(y)} \right) \right]$$

which can be thought of the amount of information gained about X from observing Y, or equivalently, information gained about Y from observing X. Note that $I(X;Y) = 0$ implies independence of the random variables since the integral is 0 if and only if

$$\ln \frac{f(x,y)}{f(x)f(y)} = 0 \implies \frac{f(x,y)}{f(x)f(y)} = 1 \implies f(x,y) = f(x)f(y).$$

Another interesting note is the relation between Kullback-Leibler Divergence and Mutual information, we can see that

$$I(X;Y) = D\left(f(x,y) || f(x)f(y)\right).$$

Some properties of mutual information are as follows:

- $I(X;Y) \geq 0$ *with equality if and only if* X *and* Y *are independent.*

- $I(X;Y) = I(Y;X)$.

- *If* F *and* G *are smooth and uniquely invertible functions then* $I\left(F(X); F(Y)\right) = I(X;Y)$.

Property (3) has rather interesting implications. The first that we cannot apply a function to a random variable X to increase the dependence between X and Y, this makes intuitive sense. The second is that mutual information is invariant to transformations of the marginal distribution of random variables. For instance, if U and V are uniformly distributed and we apply the inverse C.D.F's $F^{-1}(U) = X$ and $G^{-1}(V) = Y$ then $I(U;V) = I(X;Y)$ by this property. [1]

Linfoot's informational coefficient of correlation is defined as

$$\delta = \sqrt{1 - e^{-2I(X;Y)}}.$$

It has many of the nice properties that are desired in a dependence measure, notably it ranges from 0 to 1 and is equal to pearson's correlation coefficient if the variables follow a joint normal distribution. [4]

4.2 Copulas

A $d-$variate copula $C : [0,1]^d \rightarrow [0,1]$ is the cumulative distribution function of a random vector $\mathbf{u}(U_1, ..., U_d)$ with standard uniform margins:

$$C(\mathbf{u}) = P(U_1 \leq u_1, ..., U_d \leq u_d)$$

where $P(U_j \leq u_j) = u_j$ for $j \in 1, ..., d$ and $0 \leq u_j \leq 1$.
So, an easy example of a 2-variate copula is just the cdf of a 2 dimensional uniform distribution. But this is not the only possible copula that can be constructed for these given margins. This demonstrates the practical effectiveness of copulas for modeling different dependence structures.

Copulas also possess the following properties:

- If some component u_j is 0, then $C(\mathbf{u}) = 0$.
- $C(1, ..., 1, ..., u_j, 1, ..., 1) = u_j$ if $u_j \in [0,1]$.
- C is quasi-monotone.
- C is non-decreasing in each variable.

These properties are equivalent to stating that the copula is a cumulative distribution function with uniform marginals, and allow for any distribution function to be constructed as a copula and given marginal distribution functions, as we will see in a moment with Sklar's Theorem. [5]
Sklar's Theorem: *Every multivariate cumulative distribution function*

$$G(x_1, ..., x_d) = P(X_1 \leq x_1, ..., X_d \leq x_d)$$

of a random vector $(X_1, ..., X_d)$ can be expressed in terms of its marginals $F_i(x_i) = P(X_i \leq x_i)$ and a copula C as

$$G(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)).$$

[6]

In the case that a multivariate distribution has a density g, by the chain rule we get

$$g(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) \prod_{i=1}^{d} f(x_i)$$

where c is the density of the copula. The density of the copula can be found as $c(\mathbf{u}) = \frac{\partial C(\mathbf{u})}{\partial u_1 ... \partial u_d}$. The converse of the theorem is also true and very useful in practice. [6]

Converse of Sklar's Theorem: *Given a copula C, and marginal distributions $F_i(x_i)$ then $C(F_1(x_1), ..., F_d(x_d))$ defines a $d-$dimensional cumulative distribution function.*

The converse allows cumulative distribution functions to be constructed from a copula and desired marginal distributions. So, we can create any cumulative distribution function with the dependency structure entirely defined by a copula any univariate distribution for the marginals.
Archimedean copulas are defined by their generating function

$$\psi(t) : [0, 1] \to [0, \infty]$$

which is monotonically decreasing. The generator is said to generate a 'strict Archimedean' copula if $\lim_{t \to 0} \psi(t) = \infty$. The bivariate copula is then defined by the generator as

$$C(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2)).$$

These copulas are referred to as Archimedean. [2]

The mutual information can also be expressed in terms of copulas via Sklar's theorem, using the fact that $f(x, y) = c(F(x), F(y))f(x)f(y)$. We can then express the integral as

$$\begin{aligned}
&\int_{S_y}\int_{S_x} f(x,y)\ln\left(\frac{c(F(x),F(y))f(x)f(y)}{f(x)f(y)}\right)dxdy \\
&= \int_{S_y}\int_{S_x} f(x,y)\ln\left(c(F(x),F(y))\right)dxdy \\
&= E\left[\ln\left(c(F(x),F(y))\right)\right].
\end{aligned}$$

Recall the previous properties of Mutual Information, most notably that for smooth and uniquely invertible functions F and G then $I(X;Y) = I(F(X);G(Y))$. Let F be the c.d.f of X and let G be the c.d.f of Y, and let $U = F(X)$ and $V = G(Y)$. Then U, V are standard uniformly distributed and therefore the mutual information is

$$I(X;Y) = I(U;V) = \int_0^1\int_0^1 c(u,v)\ln\left(c(u,v)\right)dudv.$$

by Sklar's Theorem. An interesting note is that, according to this simplification, we have

$$I(U;V) = \int_0^1\int_0^1 c(u,v)\ln\left(c(u,v)\right)dudv = -H(U;V),$$

and also that

$$I(U;V) = D(c(u,v)||\pi)$$

where $c(u,v)$ is the density function for a copula and π is the independence copula i.e. π is simply a product of marginal distributions. Recall that $H(X;Y)$ is the joint-entropy of X and Y, and that $D(f||g)$ is the Kullback-Liebler divergence of two distributions.

This result is significant in that it shows the marginal distributions are inconsequential to the mutual information of a joint distribution, and that all mutual information can be determined solely from the copula. Since copulas fully describe the dependence structure of the joint distribution, it follows that mutual information fully captures all the dependence of the joint distribution, as it is a function of the copula.

4.3 The Relationship Between Kendall's τ and Linfoot's δ for Archimedean Copulas

An archimedean copula with generator $\phi(t)$ will have

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt,$$

ash shown by Genest and McKay. [3] The goal for this section is to show the relationship between τ and δ for various classes of archimedean copulas. All mutual information calculations were performed numerically using Wolfram Mathematica, up to seven significant digits.

4.3.1 The Gumbel Copula

Archimedean Copulas have the special property that they are completely defined by their generator. The generator of an Archimedean copula is function

$$\psi(t;\theta) : [0,1] \to \mathbb{R}^+.$$

If a function ψ satisfies the properties of a generator then

$$\psi^{-1}\left(\psi(u;\theta) + \psi(v;\theta)\right) = C(u,v;\theta)$$

is a 2-dimensional copula. The function

$$\psi(t;\theta) = (-\ln(t))^{\theta+1} \text{ for } \theta \in [0,\infty)$$

is the generator for the Gumbel copula. This gives the inverse generator as

$$\psi^{-1}(t) = e^{-t^{1/(\theta+1)}}.$$

The distribution function for a two dimensional Gumbel Copula is then

$$\exp\left(-\left[(-\ln(u))^{\theta+1} + (-\ln(v))^{\theta+1}\right]^{1/(\theta+1)}\right).$$

When performing a visual analysis, we can choose to examine either the CDF or the PDF of the copula. Further, we can choose a 3D plot, a contour or a scatter plot of a random sample. In the case of a Gumbel Copula with $\theta = 1.5$

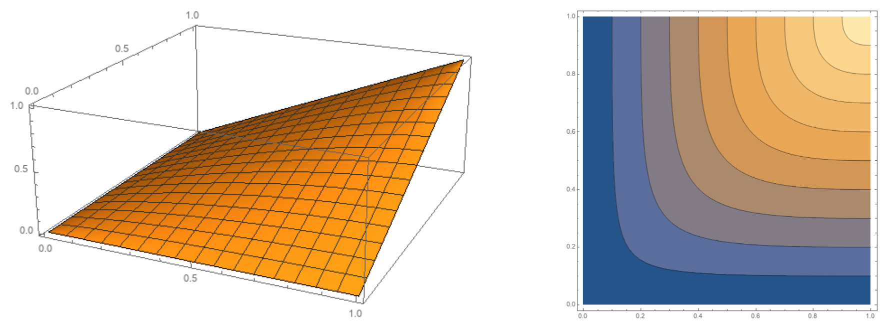

Figure 1: Above are the 3D plot and contour plot, respectively, of the CDF for Gumbel Copula with $\theta = 1.5$. Kendall's Tau is equal to .6.

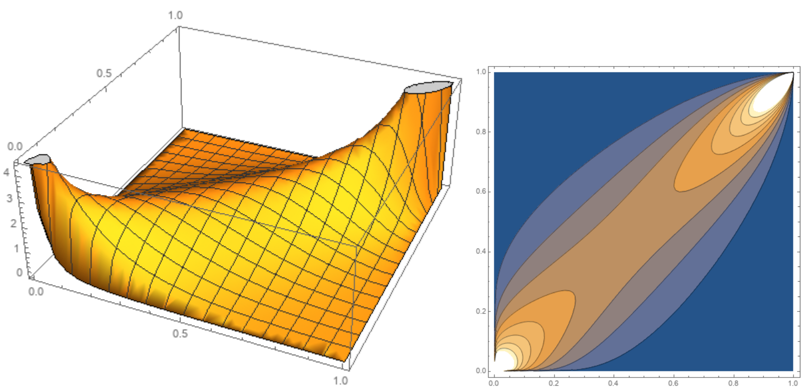

Figure 2: Above are the 3D plot and contour plot, respectively, of the PDF for Gumbel Copula with $\theta = 1.5$. Kendall's Tau is equal to .6.

the contour and 3D plot are shown. My personal preference is to examine the PDF, because it will highlight the areas of the copula which exhibit the greatest dependence. Referring to figure two, we can see that the Gumbel Copula exhibits a strong upper right tail dependence, and comparatively weaker, but still strong, lower left tail dependence, just from the plots.

Let $c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$ denote the density function for the copula. Given this, we can then calculate the mutual information of the copula via the double integral

$$I(U; V) = \int_0^1 \int_0^1 c(u, v) \ln\left(c(u, v)\right) du dv.$$

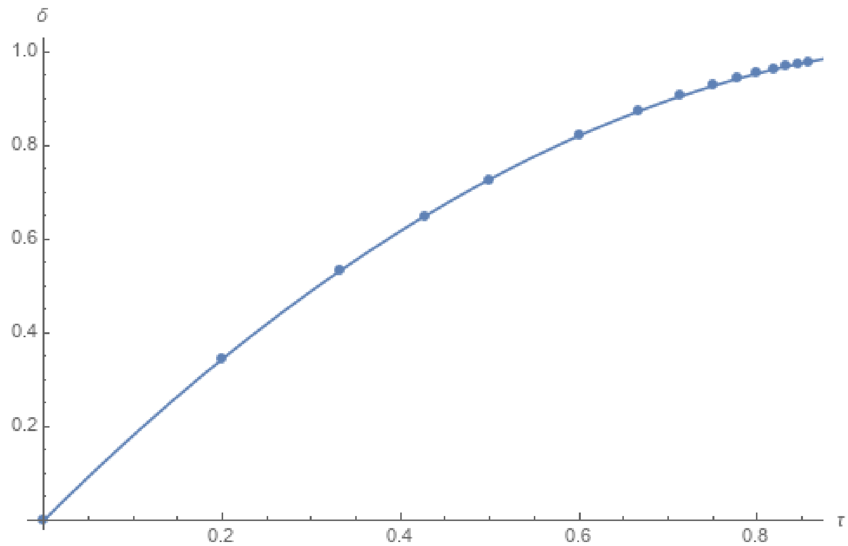

Figure 3: Above is δ as function of τ for the Gumbel Copula. The graph has been interpolated to show a continuous function.

This is a non-elementary integral, so I turned to a numerical solution.

The graph below summarizes the results. On the horizontal axis is τ, which can be calculated given a value for θ for all Archimedean copulas. In the case of the Gumbel copula, this relationship is $\tau = \frac{\theta}{\theta+1}$. On the vertical axis is the value for Linfoot's Delta, which is a function of Mutual Information. For mutual information I the Linfoot's Delta is equal to $\delta = \sqrt{1 - e^{-2I}}$. A quadratic approximation for Linfoot's delta can be found for the Gumbel copula by a simple Quadratic Regression. The equation for estimating δ is approximately

$$\hat{\delta} = 1.896\tau - .880\tau^2.$$

If we take the derivative this equation with respect to τ, we can see that the rate of information growth is high for τ close to zero and tapers off as τ increases. The first derivative of δ can be a useful tool for comparing the information growth of copulas as a function of τ.

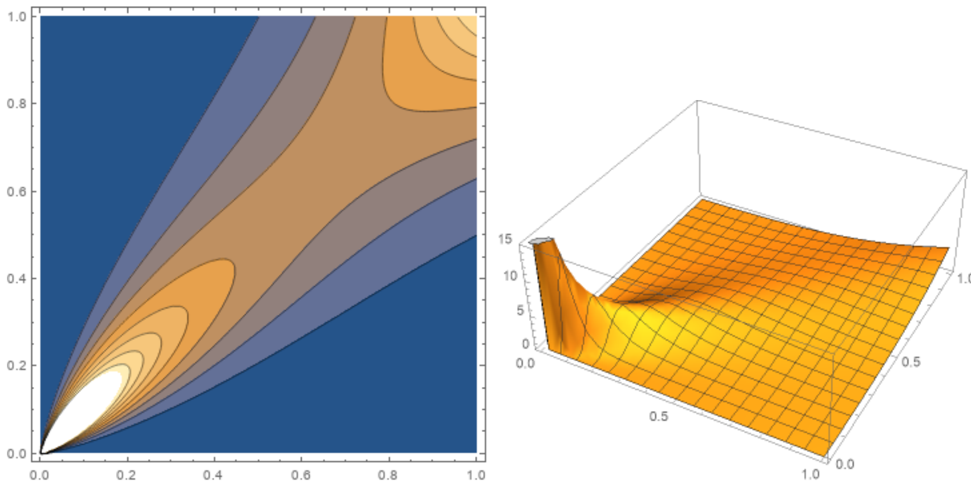

Figure 4: The contour plot, and 3D plot, respectively, for the Clayton Copula density function.

4.3.2 The Clayton Copula

The Clayton copula is defined by the generator

$$\psi(t;\theta) = \frac{(t^{-\theta} - 1)}{\theta} \text{ for } \theta \in (0, \infty).$$

This gives the inverse generator as

$$\psi^{-1}(t) = (\theta t - 1)^{\frac{-1}{\theta}}.$$

Combining the two gives us the two-dimensional Clayton Copula CDF as

$$C(u, v; \theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}.$$

The PDF can be obtained by taking the mixed partial derivative with respect to u and v. This PDF exists, but the expression in terms of elementary function, like the Gumbel Copula density function, is quite long, so I will omit it here. Instead, I will provide both a contour plot and a 3D graph of the PDF, so the copula can be visually compared to the Gumbel Copula.

In an attempt to understand the differences between the Gumbel and the Clayton copulas we can examine the differences between their contour plots. We can see that the Clayton copula exhibits strong lower tail dependence, but there is little upper tail dependence. Similarly to the Gumbel Copula, let $c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$ denote the density function for the copula. Given this, we can then calculate the mutual information of the copula via the double integral

$$I(U;V) = \int_0^1 \int_0^1 c(u,v) \ln\left(c(u,v)\right) dudv.$$

Again this integral can easily be calculated numerically.

The graph below summarizes the results. On the horizontal axis is τ, which can be calculated given a value for θ for all Archimedean copulas. In the case of the Clayton copula, this relationship is $\tau = \frac{\theta}{\theta+2}$. On the vertical axis is the value for Linfoot's Delta, which is a function of Mutual Information. For mutual information I the Linfoot's Delta is equal to $\delta = \sqrt{1 - e^{-2I}}$. A quadratic approximation for Linfoot's delta can be found for the Clayton copula by a simple quadratic regression. The equation for estimating δ is approximately

$$\hat{\delta} = 2.0505\tau - 1.0607\tau^2.$$

If we take the derivative this equation with respect to τ, we can see that the rate of information growth is high for τ close to zero and tapers off as τ increases. The first derivative of δ can be a useful tool for comparing the information growth of copulas as a function of τ. If we were to look at the differences between the derivative δ with respect to τ for the Clayton and Gumbel copulas, we can see that the Clayton copula has a higher rate of information growth relative to the Gumbel copula, up to $\tau = .427$ at which point the Gumbel copula has a higher rate of information growth.

4.3.3 The Frank Copula

The Frank copula is defined by the generator

$$\psi(t;\theta) = -\log\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right) \text{ for } \theta \neq 0.$$

This gives the inverse generator as

$$\psi^{-1}(t) = -\frac{1}{\theta}\log\left(1 + \exp(-t)(\exp(-\theta) - 1)\right).$$

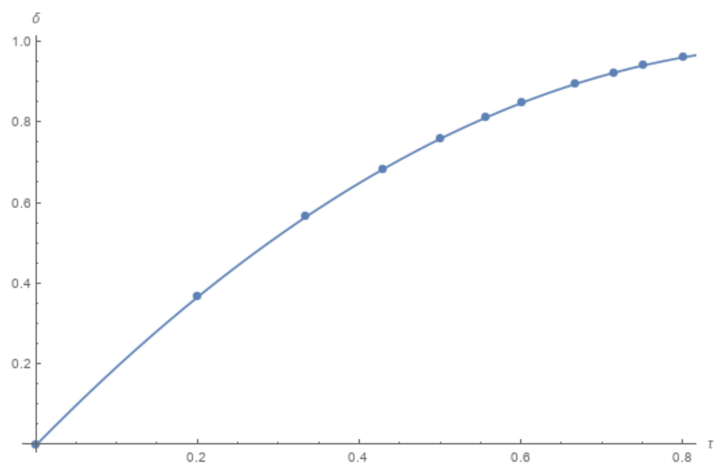

Figure 5: The best fit line for δ as a function of τ for the Clayton Copula.

Combining the two gives us the two-dimensional Frank Copula CDF as

$$C(u, v; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1) - 1}{\exp(-\theta) - 1} \right).$$

The PDF can be obtained by taking the mixed partial derivative with respect to u and v. This PDF exists, but the expression in terms of elementary function is again quite extensive and uninformative. It is useful to instead look at the contour and 3-D plots of the PDF. The Frank copula is notably symmetric, with no differences between the upper tail and lower tail. Another striking feature is the relatively constant width of the contours. There is no bowing out of the PDF near the middle, like there is in Gumbel PDF. Let $c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$ denote the density function for the copula. Given this, we can then calculate the mutual information of the copula via the double integral

$$I(U; V) = \int_0^1 \int_0^1 c(u, v) \ln (c(u, v)) \, du dv.$$

Again this integral can easily be calculated numerically.

The graph below summarizes the results. On the horizontal axis is τ, which can be calculated given a value for θ for all Archimedean copulas. In the case of the Frank copula, the expression is not as clean as the Clayton and

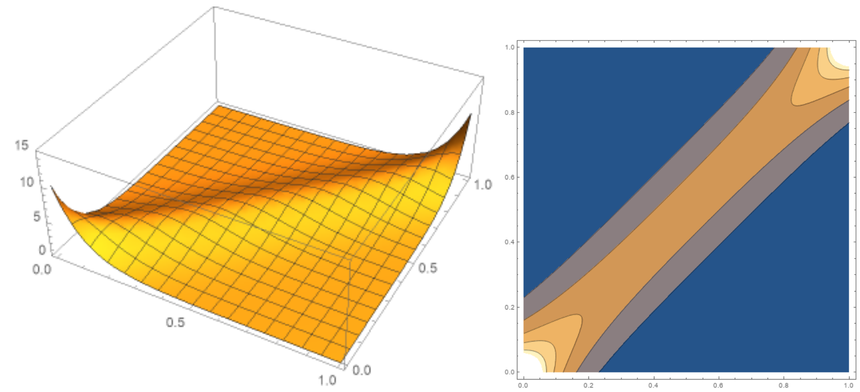

Figure 6: The 3D plot and contour plot of the Frank copula PDF.

Gumbel Copulas. The relationship is

$$1 + \frac{4\left(D_1(\theta) - 1\right)}{\theta}$$

where $D_1(\theta) = \int_0^\theta \frac{t}{\theta(e^t - 1)} dt$. On the vertical axis is the value for Linfoot's Delta, which is a function of Mutual Information. For mutual information I the Linfoot's Delta is equal to $\delta = \sqrt{1 - e^{-2I}}$. A quadratic approximation for Linfoot's delta can be found for the Frank copula by a simple quadratic regression. The equation for estimating δ is approximately

$$\hat{\delta} = 1.6776\tau - .6273\tau^2.$$

Judging from the graph we can see that this copula has the most linear relationship of the three copulas. The constant nature of the PDF supplies some reasoning to the linear nature of the relationship between τ and δ. If, as was done with the other copulas, we take the derivative of the δ function we get $\delta'(\tau) = 1.6776 - 1.2545\tau$. Because the coefficient of the τ^2 term is small relative to the other copulas, it again demonstrates the "constant-ness" of the information growth.

4.4 Applying Copulas to Foreign Exchange Data

This section explores the usability of a heuristic for fitting copulas based off the delta function explored in the previous section. The closing quotes of the EUR/USD and USD/CHF foreign exchange currency pairs were gathered for

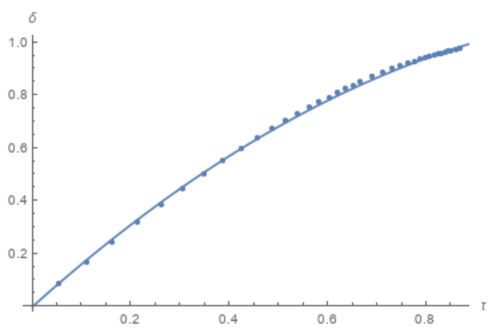

Figure 7: The best fit line for δ as a function of τ for the Frank copula.

the dates January 1, 2020-May 5, 2020. Next, the daily log-returns of the series was taken. Log returns can be calculated as

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$$

for a price p at time t. These pairs exhibit a negative correlation. Since the Joe copula is the only copula that can capture negative correlation, if we instead note this fact and take the negative log-returns for one of the pairs, we will obtain a positive correlation. For the data, the correlation statistics were $\tau = .7421$ and $\delta = .9269$. The heuristic for fitting the best copula is a simple score, S_c:

$$S_c = |\delta_c(\tau) - \delta_d|$$

where S_c is the score for a given copula, δ_c is the delta function for the copula, and δ_d is the δ of the data. The score for the Gumbel copula is .00446, the score for the Clayton copula is .01065, and the score for the Joe copula is .027354. This heuristic would give the Gumbel copula as the best fit copula for the data.

If we compare this approach to the more traditional Akaike Information Criterion approach, where the AIC is defined as

$$AIC = 2k - 2\ln(\mathcal{L})$$

where k is the number of parameters and $\mathcal{L}$ is the likelihood function for the model. The Gumbel copula model had an AIC of -256.81, the Clayton copula model had an AIC of -211, and the Frank copula model had an AIC of -231.

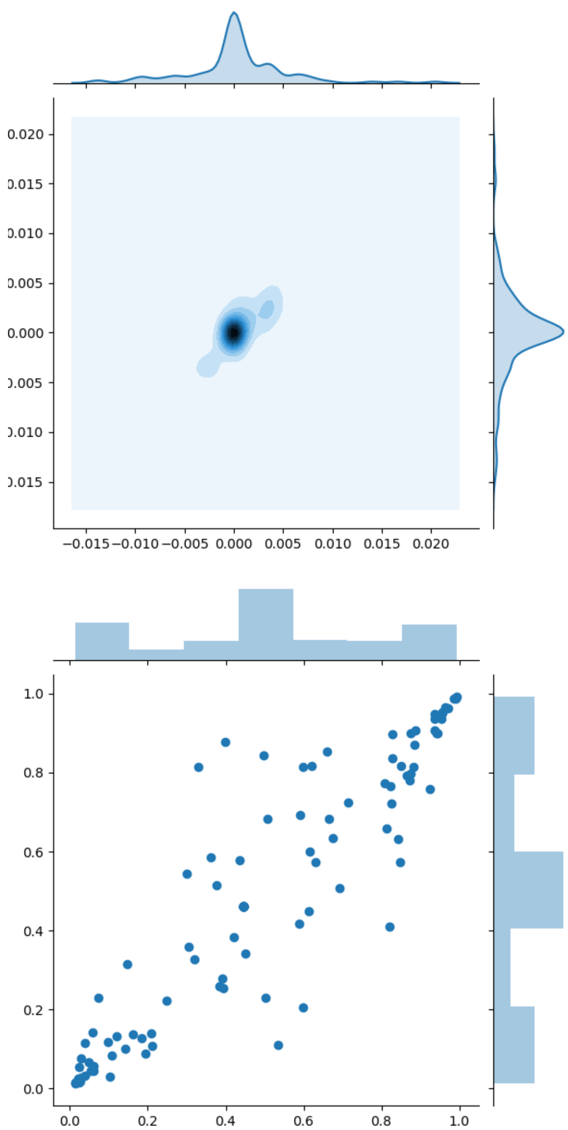

Figure 8: The joint plot for the pairs -EUR/USD and USD/CHF and the joint distribution after the marginal distributions have been made uniform.

Therefore, the heuristic agrees with the more standard method in this case. There are potential pitfalls though, as the heuristic is subject to errors in the estimation of mutual information and the regression errors from obtaining the quadratic equation. It is not hard to imagine that a dataset exists where the heuristic might disagree with the standard method using AIC.

5 References

[1] Thomas M Cover and Joy A Thomas. *Elements of information theory.* John Wiley & Sons, 2012.

[2] Christian Genest and Jock MacKay. The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283, 1986.

[3] Christian Genest and Louis-Paul Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043, 1993.

[4] Edward H Linfoot. An informational measure of correlation. *Information and control*, 1(1):85–89, 1957.

[5] Roger B Nelsen. *An introduction to copulas.* Springer Science & Business Media, 2007.

[6] Abe Sklar, A SKLAR, and CA Sklar. Fonctions de reprtition an dimensions et leursmarges. 1959.