Directional Dependence Using Concomitants of Order Statistics and Copulas

Rocherno de Jongh, Wahid Tawsif Ratul

Advisor: Professor Engin A. Sungur

University of Minnesota – Morris, Division of Science and Mathematics

Abstract

Directional dependence is a method to determine the likely casual direction of effect between two variables and is different than the direction of dependence. Directional dependence involves studying the changes in the distributional behavior of one variable not necessarily caused by another random variable. We will use order statistics and concomitant of order statistics to try and determine the directional dependence between two variables. Then, we will use a sampling algorithm to see how accurate our data is with different sample sizes. Then, using how the algorithm was set up, we will use copulas to try and get the directional dependence.

Introduction and Background

Let (X, Y) be an absolutely continuous random vector with joint cumulative distribution function, cdf, $F(x, y)$, and marginal cdfs $F_X(x)$ and $F_Y(y)$.
Let $X_1, X_2, \dots, X_n$ denote a random sample of n observations from the distribution of X. We denote ordered random variables $X_{i:n}$ by $X_{1:n}, X_{2:n}, \dots, X_{n:n}$, where
$X_{1:n} \le X_{2:n} \le \dots \le X_{n:n}$ is the order statistics of X. Let $Y_1, Y_2, \dots, Y_n$ denote a random sample of n observations from the distribution of Y. The value of X associated with $X_{i:n}$ is called the concomitant of the ith ordered statistics and will be represented as $Y_{[i:n]}$.

Methods

Our model for understanding directional dependence through order is based on, that if the directional dependence from X to Y is strong then X will initiate a "substantial" change in the ordering of Y reflected on the concomitants leading to a weaker dependence between concomitant and the marginal ordering of Y. Strength of change imposed (weakness of dependence) may be used to determine the dominant directional dependence. This is graphically represented by figure 1. Sungur argued that $X \rightarrow Y$ iff $\left|r_{Y_{[i:n]}Y_{i:n}}\right| < \left|r_{X_{[i:n]}X_{i:n}}\right|$.
We will use a bivariate sampling algorithm to see how our results changes as the sample size changes. The algorithm works as following:
Let f_X, f_Y, be univariate p.d.f. We will sample from a joint p.d.f. with given marginals f_X and f_Y. Let $n \ge 2$, let σ be a permutation on $\{1, \dots, n\}$, and let $0 \le p \le 1$.

1. Sample independent deviates $X_1, \dots, X_n$ from f_X.
2. Sample independent deviates $Y_1, \dots, Y_n$ from f_Y.
3. Let $X_{(i)}, Y_{(j)}, 1 \le i, j \le n$, denote the order statistics. Then the final bivariate sampled vector is (U, V) with

$$(U,V) = \begin{cases} (X_1, Y_1) & \text{with probability } p \\ \left(X_{(1)}, Y_{(\sigma(1))}\right) & \text{with probability } \dfrac{1-p}{n} \\ \left(X_{(2)}, Y_{(\sigma(2))}\right) & \text{with probability } \dfrac{1-p}{n} \\ \vdots & \\ \left(X_{(n)}, Y_{(\sigma(n))}\right) & \text{with probability } \dfrac{1-p}{n} \end{cases}$$

Acknowledgement

We would like to give a special thank to Professor Engin A. Sungur for invaluable mentorship throughout the summer.

We will use the way that the sampling algorithm is set up to create copulas. Copulas eliminate the influence of marginal behavior and provides a clear look at dependence structure. Sklar (1973) shows that any multivariate distribution of its marginals, say F, can be represented as a function of its marginals, say F_X and F_Y, by using a copula C, i.e., $F(x, y) = C(F_X(x), F_Y(y))$. In simple terms, copula is a bivariate distribution function of two uniform random variables. Directional dependence is a property of joint distribution, i.e. copula. We are interested in asymmetric copulas, these can tell us about the directional dependence of the variables in joint behavior.

Application Distribution Cycle

We will now apply the distribution cycle to a real world data. We will use a data set which has as X = Percent of adults aged 18 years and older with obesity, and Y = Percentage of adults with diabetes, where n = 52. We can see the distribution for this data in figure 1. Using our sampling algorithm on this data,

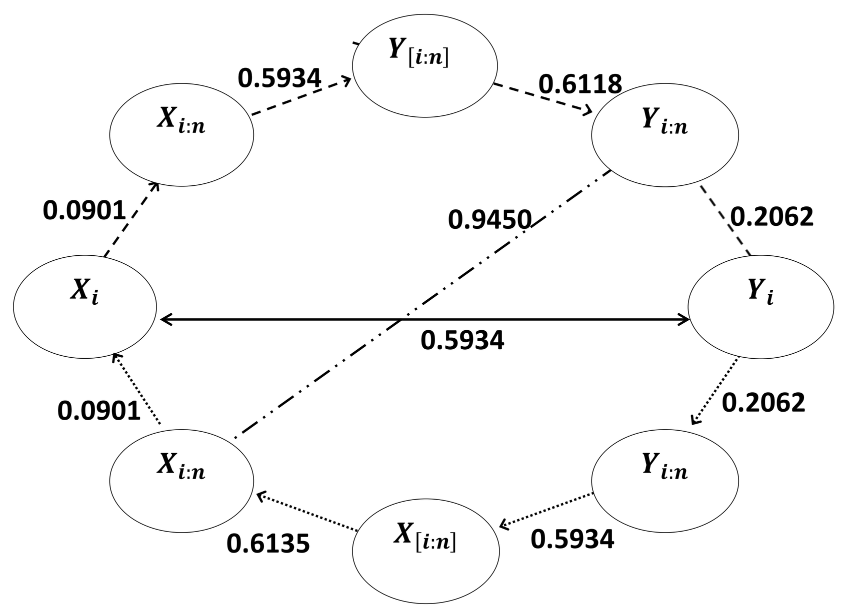

Figure 1: Distribution Cycle of X = Percentage of adults with obesity, and Y = Percentage of adults with diabetes

Application Copulas

From the algorithm, we can get the marginals, $F_U(u)$ and $F_V(v)$, and also the joint distribution, $F_{UV}(u, v)$. We will do it for n = 2. We get $P(U \le u) = F_U(u) = pF_X(u) + \frac{1-p}{2}\left[\left(1-\left(1-F_X(u)\right)^2\right) + \left(F_X(u)\right)^2\right]$, we can get $F_V(v)$ in similar way. Because of this, we will have two possible joint distributions,
$F_{UV}(u,v) = pF_X(u)F_Y(v) + \frac{1-p}{2}\left[\left(1-\left(1-F_X(u)\right)^2\right)\left(1-\left(1-F_Y(v)\right)^2\right) + \left(F_X(u)\right)^2\left(F_Y(v)\right)^2\right]$ (1) or
$F_{UV}(u,v) = pF_X(u)F_Y(v) + \frac{1-p}{2}\left[\left(1-\left(1-F_X(u)\right)^2\right)\left(F_Y(v)\right)^2 + \left(F_X(u)\right)^2\left(1-\left(1-F_Y(v)\right)^2\right)\right]$ (2).
From here, we get the copula $C(u, v) = uv[(p-1)(1-u)(1-v)+1]$.

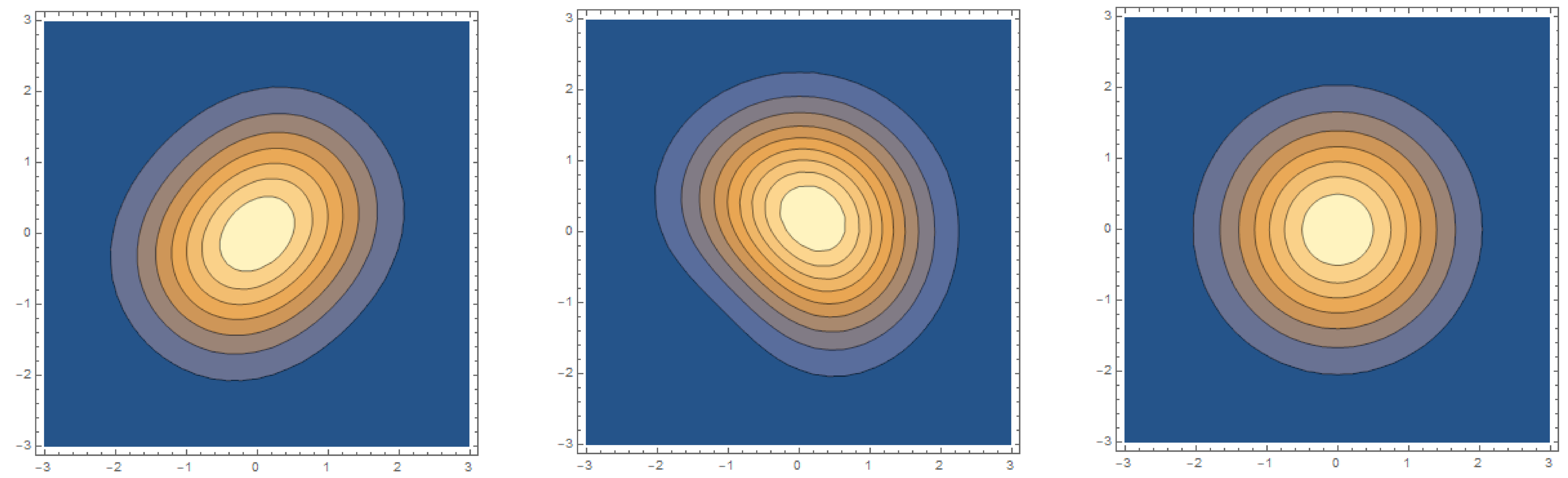

Figure 2: Contour plot of density function of (1) with p is 0.5, contour plot of density function of (2) with p is 0.5, contour plot of density function of (1) and (2) with p is 0.

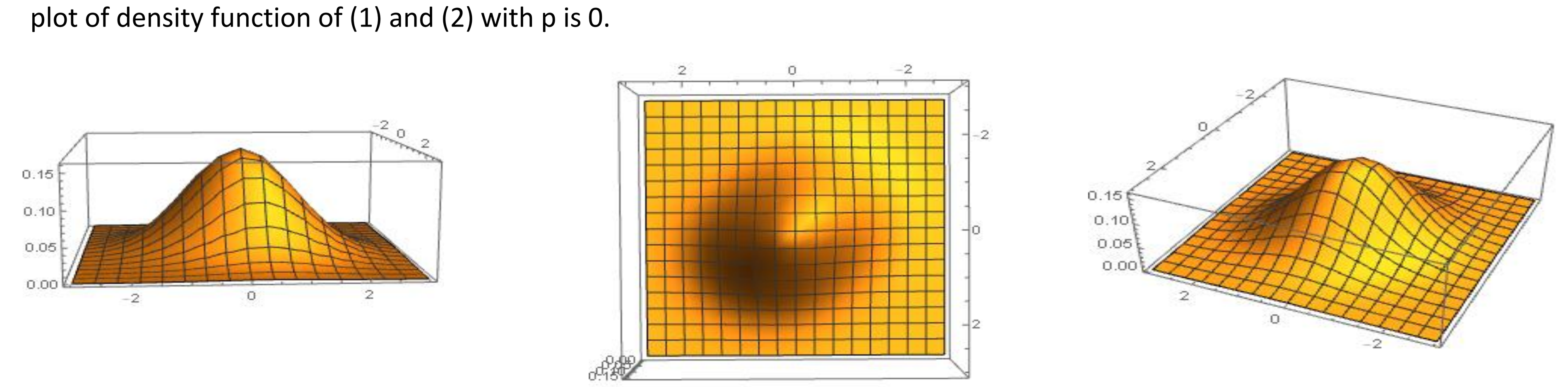

Figure 3: 3D plots of density function of (1), with p is 0.5.

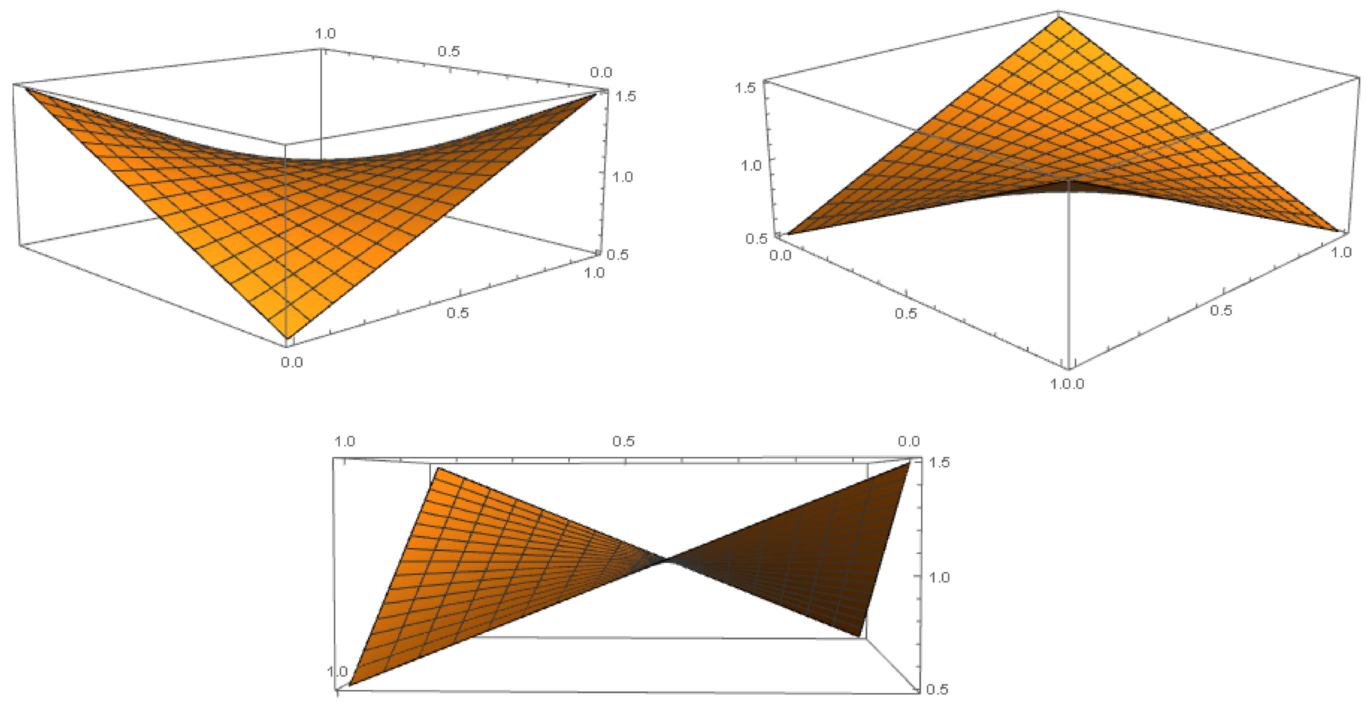

Figure 4: Copula Density of (2), when p is 0.5.

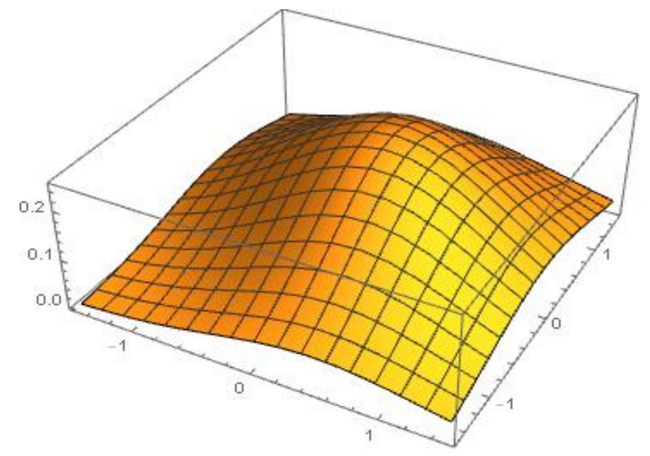

Figure 5: 3D plots of density function of (2) with p is 0.5.

Results

From the Distribution Cycle we can see that 0.6618 < 0.6135, so percent of adults with obesity affects percent of adults with diabetes. Using our sampling algorithm, we get that $\left|r_{Y_{[i:n]}Y_{i:n}}\right| < \left|r_{X_{[i:n]}X_{i:n}}\right|$ is true in about 32% of the times, but this is probably because our number of observations is small.
The contour plot of the density function of (2) is a bit different from the contour plot of the density function of (1) when p is equal to 0.5, but we can see that the contour plot of the density functions of (1) and (2) are both the same, and that is expected looking at the equations. We see that we get a symmetric copula when n = 2, so there is not much we can say of the directional dependence. We have the preliminary work done for the copula for when n = 3, but because of lack of space and also we are not totally sure if the copula is correct yet. But, we have some confidence that when n = 3, we will get an asymmetric copula.

Current and Future Work

We are continuing our work on finding copulas using the setup of the algorithm. What wish to accomplish is to generalize what type of copula we will find no matter the size of n. A couple of difficulties we will encounter is that because there are $n!$ possible joint distribution function for each n, so the number of joint distribution gets really big even for relatively small n. Also, when n gets larger, we get really big power of polynomials.

References

Granville, Vincent. *Sampling from a bivariate distribution with known marginals*. National Institute of Statistical Sciences, 1996.
Sungur, Engin A. "*A note on directional dependence in regression setting.*" Communications in Statistics—Theory and Methods 34.9-10 (2005): 1957-1965.
Sungur, Engin A. "*Directional Dependence and Concomitants of Order Statistics.*" Unpublished.