**Modeling Bivariate Directional Dependence using Order Statistics and Concomitants**

**Wahid Tawsif Ratul**

**University of Minnesota, Morris**

In this paper, we introduce a new algorithm to sample bivariate directional dependence based on the applications of order statistics and concomitants. We randomly simulated the data using Normal and Uniform Probability Distributions and applied the methods of concomitants that led us to the conclusion that variables that are Uniformly Distributed tend to show the strongest directional dependency. Further simulations were done using the Uniform Distribution and the randomly generated algorithm used to model directional dependence led us to develop various combinations of model parameters that would help identify the strongest bivariate directional dependency. The setup introduced in the research can be easily adapted to a wide range of dependence modeling of the same structure.

**Keywords:** Concomitants, Order Statistics, Directional Dependence, Correlation

## 1. Introduction and motivation (progression of research)

The main goal of this research is to carry out a simulation study that uses the fundaments of order statistics and concomitants to exhibit bivariate directional dependence property. Our investigation started with looking at the Normal and Uniform Probability Distributions to randomly simulate two sets of variables (X, Y) and identify the suitable probability distribution that exhibits strongest directional dependence properties that will be used for further simulations of model parameters on the latter stages of the research.

### *1.1 Directional Dependence*

Directional Dependence is a concept that can be defined in various ways and is different than the direction of dependence. In a bivariate setting, if the large and small values of the random variables tend to occur together then the direction of dependence is positive; and if the large values of one variable tend to occur with small values of the other then the direction of dependence is negative, *Nelsen and Úbeda-Flores (2011)*. On the other hand, directional dependence also involves studying changes in distributional behavior of one variable initiated not necessarily caused by another random variable. Therefore, it is different than studying causality which can only be set up through carefully designed experiments. Directional Dependency can be exhibited in numerous ways as suggested by *Dodge and Rousson (2000), Muddapur (2005), and Sungur (2005)* where they all have studied in regression setting. In the first two, they suggested an approach to decide about the direction of the regression line. Their approach is based on the use of coefficients of skewness of the two random variables. It has been shown that under the assumption that the error term is symmetric the cube of the Pearson's correlation can be calculated as the ratio of these two skewness measures. *Dodge and Rousson (2000)* claimed that the deviation from symmetry can be used to assess the directional dependence and the response variable will always have less skewness than the explanatory variable. *Sungur (2005)* argued that the conditional not the marginal behavior of the variables will determine the dependence structure and suggested the use of copula approach. In this case, existence of directional dependence in regression setting implies a different copula regression function in two directions and requires nonsymmetric copula. *Sungur and Celebioglu (2011)* discussed generating a class of copulas with directional dependence property by introducing direction covariate and set up the connection between directed graphs which is essentially the primary article used as a guide for the entire research.

#### *1.2 Order Statistics and Concomitants*

One possible way to understand the changes in distribution of one variable initiated by another is to study the concomitants of order statistics. 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 $F_Y(y)$. Consider random sample of *n* observations from the distribution of *X* and *Y*, and let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics of *X*. The value of *Y* associated with the $X_{i:n}$ is called the concomitant of the $i^{th}$ ordered statistics of the *X* and will be represented by $Y_{[i:n]}$. Similarly, $X_{[i:n]}$ is the X-concomitant of the $Y_{i:n}$. A function of $Y_{i:n}$ and $Y_{[i:n]}$ ($X_{i:n}$ and $X_{[i:n]}$) can be considered as an estimator of the population correlation. In this direction *Schechtman and Yitzhaki (1987)* have proposed sample Gini correlation coefficient for the bivariate normal case. *Barnett et all (1976)* studied the estimation of the correlation coefficient in a bivariate normal distribution using only the information contained in the concomitants of the ordered observations of one of the variables. The joint distribution of $\left(Y_{i:n}, Y_{[j:n]}\right)$ is obtained by Tsukibayashi (1998) for the case $i = j = n$, and by *He and Nagaraja (2009) f*or all $i, j = 1,2, \dots, n$. Also, *He and Nagaraja (2009)* have proposed a family of estimators of correlation between *X* and *Y* using *Y*-concomitants of *X*-order statistics and *Y*-order statistics from bivariate normal samples. In this paper we will investigate the use of $Y_{i:n}$ and $Y_{[i:n]}$ ($X_{i:n}$ and $X_{[i:n]}$) in determination of directional dependence, construct an order-based measure of directional dependence using the sampling linear algorithm discussed by *Sungur and Celebioglu (2011)* to construct asymmetric copulas that exhibits directional dependence properties.

## 2. Order Statistics and Directional Dependence

#### *2.1 Directional Dependence through order*

Directional dependence can be introduced to a random phenomenon in different ways. For example, it may be due to the inherited characteristics of the variables introduced by time and space. It also may be introduced by the restrictions on the realization of one variable imposed by the other variable. In this paper, we will look at $Y_{[i:n]}$ ($X_{[i:n]}$), ordering on *Y* (*X*) introduced by ordering in *X* (*Y*), and compare it with the $Y_{i:n}$ ($X_{i:n}$), marginal ordering on *Y* (*X*). Let $\boldsymbol{X} = (X_1, \dots, X_n)^T$and $\boldsymbol{Y} = (Y_1, \dots, Y_n)^T$; $\boldsymbol{X}_{+:n} = (X_{1:n}, \dots, X_{n:n})^T = \{X_{i:n}, j = 1,2 \dots, n\}^T$ and $\boldsymbol{Y}_{+:n} = (Y_{1:n}, \dots, Y_{n:n})^T = \{Y_{i:n}, j = 1,2 \dots, n\}^T$; $\boldsymbol{X}_{[+:n]} = \left(X_{[1:n]}, \dots, X_{[n:n]}\right)^T = \left\{X_{[i:n]}, j = 1,2 \dots, n\right\}^T$and $\boldsymbol{Y}_{[+:n]} = \left(Y_{[1:n]}, \dots, Y_{[n:n]}\right)^T = \left\{Y_{[i:n]}, j = 1,2 \dots, n\right\}^T$ be random vectors of observations, order statistics, and concomitants, respectively. Each one of these random vectors provides invaluable information on the underlying dependence structure. For example, the $Y_{[j:n]}$ are not necessarily ordered, but can be expected to reflect the association between tending to lead to the values of $Y_{[j:n]}$ in roughly ascending order, and , similarly negative association to $Y_{[i:n]}$ in descending order (see Barnett et al., 1976).

### *2.2 Distribution Cycle*

Figure 1, that we will refer as the dependence cycle, gives six of the key variables in our study and their roles on understanding the distribution of the random pair $(X, Y)$. The pairwise relationships between these variables provide various information such as marginal trend, equality of the marginal distributions, (unidirectional) dependence, and directional dependence. Stronger dependence on original data on *X*(*Y*) and order statistics of it will indicate an upward or downward trend in the data. Studying original data pairs on *X* and *Y,* and order statistics and their concomitants will inform us the strength of dependence but not the directional dependence. Also, the equality of the two marginal distributions will reflect on the strength of dependence between associated order statistics.

Our model for understanding directional dependence through order is based on the following argument as it is graphically represented in Figures 1 and 2. If the directional dependence from $X \rightarrow Y > Y \rightarrow X$ then *X*(*Y*) will initiate a substantial change in the ordering of *Y* (*X*) reflected on the concomitants leading to a weaker dependence between concomitant and the marginal ordering of *Y*(*X*)*.* Strength of change imposed (weakness of dependence) may be used to determine the dominant directional dependence.

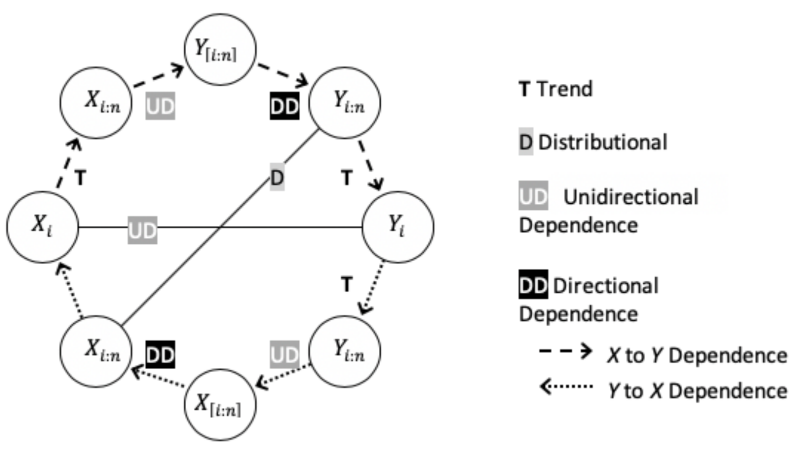


**Figure 1. Distribution Cycle.** The key random variables and their roles on understanding marginal and joint behaviors of $(X, Y)$.

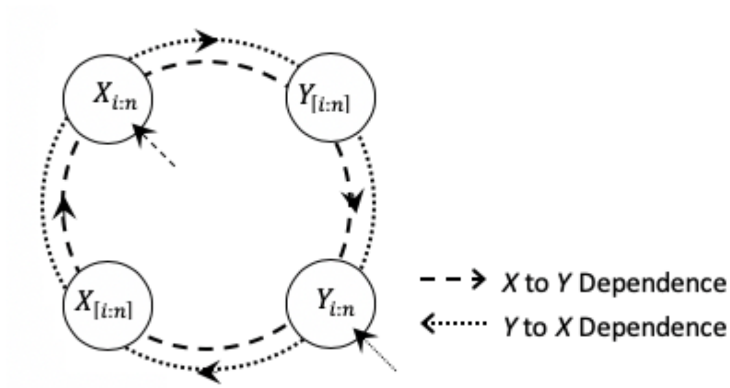


**Figure 2. Dependence Cycle.** The key random variables and their roles on understanding dependence structure of $(X, Y)$.

We can state our objective in simple terms as measuring the change in the natural ordering of *Y* induced by the ordering of *X* through its concomitant. It can be argued if *X* is affecting *Y* through ordering it will lead to a weaker relationship between the order statistics of *Y* and the *Y* concomitant of the *X* ( $Y_{[i:n]}$). Under this objective the key components that will concentrate on are $X_{i:n}, X_{[i:n]}, Y_{(i:n)}, Y_{[i:n]}$ as $r_{XY}$ is explained as the sample linear correlation coefficient. It measures the strength and direction of linear association between two quantitative variables and are the following:

$$r_{XY} = \frac{1}{(n-1)s_X s_Y} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \text{ ----- (1)}$$

$$r_{X\to Y} = r_{Y_{:n}Y_{[:n]}} = \frac{1}{(n-1)s_Y^2} \sum_{i=1}^{n} (y_{i:n} - \bar{y})\left(y_{[i:n]} - \bar{y}\right) \text{ ----- (2)}$$

$$r_{Y\to X} = r_{X_{:n}X_{[:n]}} = \frac{1}{(n-1)s_X^2} \sum_{i=1}^{n} (x_{i:n} - \bar{x})\left(x_{[i:n]} - \bar{x}\right) \text{ ----- (3)}$$

where, $s_X^2$ and $s_Y^2$ are the sample variances for *X* and *Y*, respectively. Also, will consider the case that marginal distributions of *X* and *Y* are known and are the following:

$$R_{XY} = \frac{1}{n\sigma_X\sigma_Y} \sum_{i=1}^{n} (Y_i - \mu_Y)(X_i - \mu_X) \text{ ----- (4)}$$

$$R_{X\to Y} = R_{Y_{:n}Y_{[:n]}} = \frac{1}{n\sigma_Y^2} \sum_{i=1}^{n} (Y_{i:n} - \mu_Y)\left(Y_{[i:n]} - \mu_Y\right) \text{ ----- (5)}$$

$$R_{Y\to X} = R_{X_{:n}X_{[:n]}} = \frac{1}{n\sigma_X^2} \sum_{i=1}^{n} (X_{i:n} - \mu_X)\left(X_{[i:n]} - \mu_X\right) \text{ ----- (6)}$$

where, $\mu_X$ and $\sigma_X^2$ ($\mu_Y$ and $\sigma_Y^2$) are the population mean and variance for *X* (*Y*), respectively. Note that, $E[R_{XY}] = \rho_{XY}$.

He and Nagaraja (2007) suggested the following to estimate the directional dependence between two variables:

$$r_{Y_{:n}Y_{[:n]}} = \frac{\sum_{i=1}^{n} (y_{i:n} - \bar{y})\left(y_{[i:n]} - \bar{y}\right)}{\sum_{i=1}^{n} (y_{i:n} - \bar{y})^2} \text{ ----- (7) as an estimator of } \rho_{X\to Y}.$$

$$r_{X_{:n}X_{[:n]}} = \frac{\sum_{i=1}^{n} (X_{i:n} - \bar{X})\left(X_{[i:n]} - \bar{X}\right)}{\sum_{i=1}^{n} (X_{i:n} - \bar{X})^2} \text{ ----- (8) as an estimator of } \rho_{Y\to X}$$

### *2.3 Introduction to sampling algorithm*

Our next approach to modeling directional dependence required us to focus on the mathematical approach towards building linear combinations of independently distributed random variables as demonstrated by *Sungur and Celebioglu (2011).* Let us denote the transformed random pair with $(X, Y)$ are the linear combinations of the independent normally distributed or uniform pair with $(U, V)$. Along with the positive semi-definite matrix of the form $\begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}$ such that $\alpha_1 > \alpha_2, \beta_1 \leq -\beta_2,\ \alpha_1, \alpha_2, \beta_1 > 0, \beta_2 < 0)$ have a common bivariate distribution and (X,Y) that can be expressed in the form:

$$X = \alpha_1 * U + \ \alpha_2 * V \text{ ----- (9)}$$
$$Y = \beta_1 * U + \ \beta_2 * V \text{ ----- (10)}$$

Suppose that a list of $(U_i, V_j)$ where $i\ =\ 1{,}2, \ldots$, n are *n* independent randomly generated pairs having a common bivariate distribution of $U_i, V_j$ along with the positive semi-definite matrix mentioned earlier can be expressed in the form:

$$X_i = \alpha_1 * U_i + \ \alpha_2 * V_i \text{ ----- (11)}$$
$$Y_i = \beta_1 * U_i + \ \beta_2 * V_i \text{ ----- (12)}$$

### *2.4 Normal Distribution and Directional Dependence cycle*

Once the independent set of $(X_i, Y_i)$ are randomly generated by normally distributed $(U_i, V_i)$ we were able to create the key columns of $X_{i:n}, X_{[i:n]}, Y_{(i:n)}, Y_{[i:n]}$ to examine directional dependence using the methods mentioned earlier. The following directional dependence cycle gives us an extensive relationship between each of the components of $(X, Y)$.

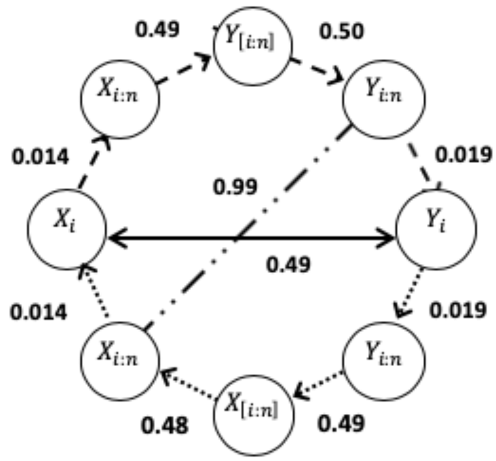


**Figure 3.** Distributional circle showing the pairwise correlations for X & Y’s generated from Normal Distribution.

As mentioned earlier, the key elements required to understand the dependence structure of (X,Y) are the $r_{X\to Y}$ and $r_{Y\to X}$ which are basically the correlations of $r_{Y_{:n}Y_{[:n]}}$ and $r_{X_{:n}X_{[:n]}}$. The key components to pay attention from the estimation visualized on the directional dependence cycle are $r_{Y_{[i:n]}Y_{i:n}} = 0.50$ and $r_{X_{[i:n]}X_{i:n}} = 0.48$.
Therefore, we suggest using the following criteria for determining the dominant direction of dependence in terms of ordering:

**Definition 1.** We will say that $X \to Y$ is the dominant direction of dependence in terms of ordering, if and only if $\left|r_{Y_{[i:n]}Y_{i:n}}\right| < \left|r_{X_{[i:n]}X_{i:n}}\right|$ or $\left(\left|R_{Y_{[i:n]}Y_{i:n}}\right| < \left|R_{X_{[i:n]}X_{i:n}}\right|\right)$. From the Normal Distribution Cycle (figure 3) we observed that $\left|r_{Y_{[i:n]}Y_{i:n}}\right| \approx \left|r_{X_{[i:n]}X_{i:n}}\right|$. Hence in this case, it is very challenging to conclude variables (X,Y) have any directional dependence.

### *2.5 Uniform Distribution and Directional Dependence cycle*

Once the independent set of $(X_i, Y_i)$ are randomly generated by uniformly distributed $(U_i, V_i)$, we were able to demonstrate directional dependence as explained in the following figure 4:

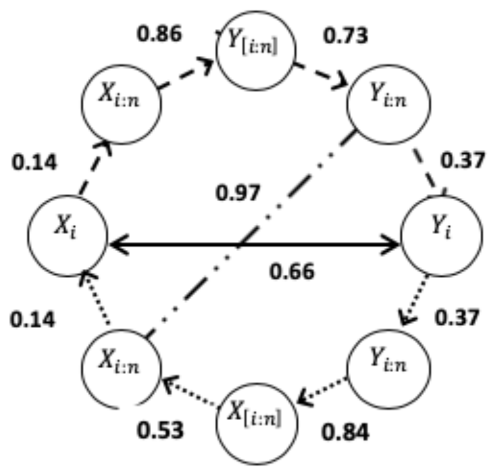


**Figure 4.** Distributional circle showing the pairwise correlations for X & Y's generated from Uniform Distribution.

The key components to pay attention from the estimation visualized on the directional dependence cycle are $r_{Y_{[i:n]}Y_{i:n}} = 0.73$ and $r_{X_{[i:n]}X_{i:n}} = 0.53$. From the Uniform Distribution Cycle (figure 4) we observed that $\left|r_{Y_{[i:n]}Y_{i:n}}\right| > \left|r_{X_{[i:n]}X_{i:n}}\right|$. Hence in this case, there is very strong evidence that $Y(x)$ causes a substantial change in marginal distribution as compared to $X(y)$ or in other words $X \to Y$.

## 3. Simulating Random parameters

### *3.1 Extension of the sampling algorithm*

The final extension of the research would be to capture the strongest directional dependence between the independent set of $(X, Y)$. The model parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ are randomly simulated using uniform distribution abiding by the conditions mentioned earlier in the making of the linear models (9) and (10) and are in the form:

$$\begin{bmatrix} U_1 & V_1 \\ \vdots & \vdots \\ U_n & V_n \end{bmatrix} * \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_n & Y_n \end{bmatrix} \text{ ----- (13)}$$

Using the model (13) we generated 100 combinations of parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ that satisfied the conditions to generate 100 list of $(X, Y)$. Each set of $(X_i, Y_i)$ were used to incorporate the key components of order statistics and concomitants $(X_{i:n}, X_{[i:n]}, Y_{(i:n)}, Y_{[i:n]})$. The key elements required to understand the dependence structure of (X,Y) are the $r_{X\to Y}$ and $r_{Y\to X}$ which are basically the correlation of $r_{Y_{:n}Y_{[:n]}}$ and $r_{X_{:n}X_{[:n]}}$. Strength of Directional Dependence can be computed as $\Delta r = r_{X\to Y} - r_{Y\to X}$. Our final objective was to find the different combinations of parameters that would exhibit the strongest directional dependence and the following section would illustrate the different combinations of parameters in the sampling algorithm that exhibits the strongest directional dependence and how it changes as the parameters fluctuate.

### *3.2 Directional Dependence and parameters table*

As mentioned earlier in the previous section, the list of combinations of model parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ were generated using the uniform distribution that helped us create a list of 100 data frames consisting of columns of $X_{i:n}, X_{[i:n]}, Y_{(i:n)}, Y_{[i:n]}$. For every list of data frame, we measured the $\Delta\alpha = \alpha_1 - \alpha_2$, $\Delta\beta = \beta_1 - \beta_2$ and $\Delta r = r_{X\to Y} - r_{Y\to X}$ as illustrated in the following table:

| $\Delta\boldsymbol{\alpha}$ | $\Delta\boldsymbol{\beta}$ | $\boldsymbol{r_{X\to Y}}$ | $\boldsymbol{r_{Y\to X}}$ | $\Delta\boldsymbol{r}$ |
|---|---|---|---|---|
| 4.06 | 29.12 | -0.31 | -0.30 | -0.01 |
| 22.4 | 51.28 | 0.37 | 0.31 | 0.06 |
| 12.51 | 16.75 | 0.05 | 0.06 | -0.01 |
| 7.18 | 30.48 | 0.01 | 0.00 | 0.01 |
| 0.26 | 28.08 | 0.23 | 0.22 | 0.01 |
| 10.43 | 20.58 | 0.75 | 0.70 | 0.05 |

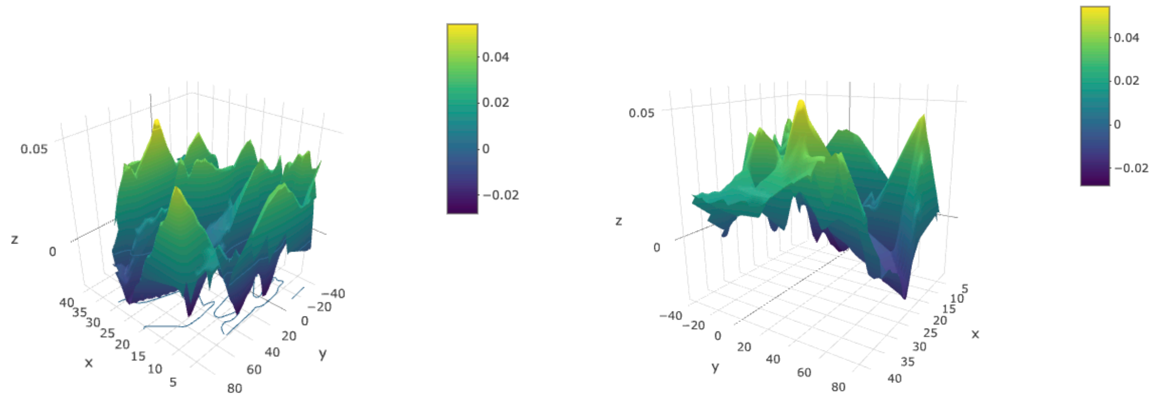

**Figure 4**. 3D Surface Plots at different angles showing Directional Dependence with different combinations of parameters where ($X$=$\alpha_1 - \alpha_2$, $Y$=$\beta_1 - \beta_2$, $Z$= $r_{X\to Y} - r_{Y\to X}$).

If we look at the surface plot figure 4, we can easily see the directional dependence structure. The plot exhibits the strongest directional dependence (the peaks) for various values of $\alpha_1, \alpha_2, \beta_1, \beta_2$. The strongest directional dependence is shown at $\Delta\alpha = 4.32,\ \Delta\beta = -11.98,\ \Delta r = 0.06$). However, there are certain combinations of the parameters that may not or partially exhibit directional dependence property. Thus, the next section examines certain cases for $\alpha_1, \alpha_2, \beta_1, \beta_2$ for $(X_i, Y_i)$.

## 4. Discussion

### *4.1 Relationship between model parameters and Directional Dependence*

One of the key things to focus on is the changes in the models parameters $\Delta\alpha,\ \Delta\beta$ affect directional dependence between $(X, Y)$**.** In order to examine the possible research question, the following is a scatterplot that will help us explain any relationship that exists (if any):

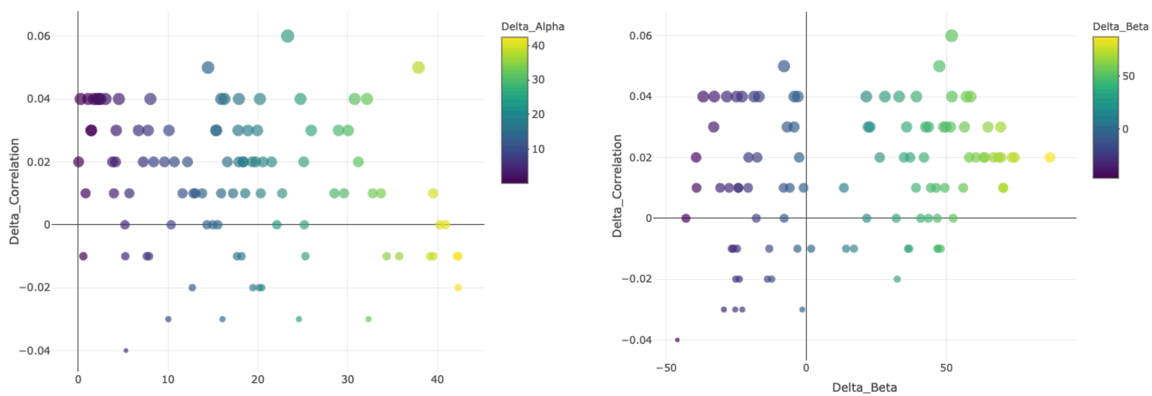

**Figure 5**. Scatterplots explaining the relationship between $\Delta\alpha$ and $\Delta\beta$ on the horizontal axes and $\Delta r$ on the vertical axes.

From the figure 5 above we can tell that $\Delta\alpha$ and $\Delta\beta$ have no concrete relationship with $\Delta r$. There are very few specific cases where the directional dependence is strongest when $\Delta\alpha = 22.4$ and $\Delta\beta = 51.28$. There are several cases where $(X, Y)$ exhibits no directional dependence ( $r_{X\to Y} - r_{Y\to X} = 0$ ) at several set of points for $\Delta\alpha$ $and$ $\Delta\beta$.

#### *4.2 Directional Dependence when variables have different correlations*

The last stage of the research is to look into cases that shows directional dependence when the uniform randomly generated $(U, V)$ exhibits characteristics of stronger correlation ($0.7 < \rho_{XY} < 1$) and no correlation ($\rho_{XY} = 0$) between variables $(X, Y)$. The following scatterplots will explain if any relationship exists between the directional dependence and changes in model parameters when the randomly generated $(X, Y)$ are strongly correlated ($\rho_{XY} = 0.9$).

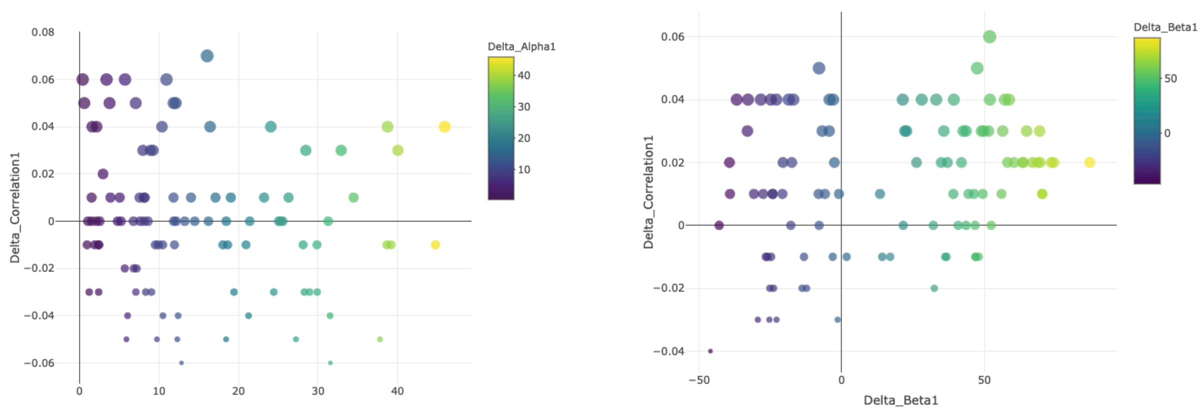


**Figure 6**. Scatterplots explaining the relationship between $\Delta\alpha$ and $\Delta\beta$ on the horizontal axes and $\Delta r$ on the vertical axes when $(X, Y)$ are strongly correlated.

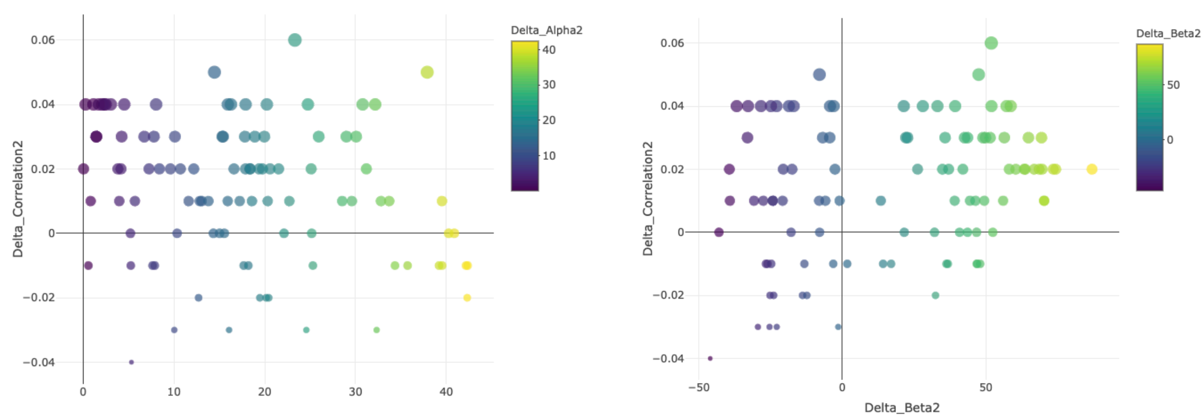

**Figure 7**. Scatterplots explaining the relationship between $\Delta\alpha$ and $\Delta\beta$ on the horizontal axes and $\Delta r$ on the vertical axes when $(X, Y)$ are not correlated.

By looking carefully at the scatterplots (figure 6 and 7), we can tell that correlation behavior between $(X, Y)$ has no effect on the directional dependence between $(X, Y)$ as the scatterplots suggest. The random simulation suggests directional dependence ( $r_{X\rightarrow Y} - r_{Y\rightarrow X}$) are very similar at different combinations of model parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ when $(X, Y)$ are both strongly and weakly correlated.

### *4.3 Future work*

The sample algorithm (9) and (10) introduced by *Sungur and Celebioglu (2011)* and in this paper possess directional dependence property and the use of directional variable in discrete and continuous forms will give researches a chance to create more general dependence models that satisfies the conditions of the model parameters that are uniformly distributed. The dependence structure could change based on the introduction of the third variable which might be discrete and continuous. Selection of a discrete or continuous direction variable will heavily depend on the researcher's interest. In a discrete setup the Uniform direction variable has an advantage of using order statistics and concomitants techniques to layout dependence structure as demonstrated in this paper. Increasing the dimensions of model parameters, $\alpha_1, \alpha_2, \beta_1, \beta_2$, and generating $(U, V)$ with different probability distribution will be a research interest in future. Moreover, it will be very interesting to see how the dependence structure behaves if $\alpha_1, \alpha_2$ are kept constant with varying $\beta_1, \beta_2$ and vice versa. The sampling algorithm given in (11) and (12) bases on the linear combinations of independent uniform variables has promising features to achieve the objective.

## 5. References


[1] Sungur, Engin A., and Salih CelebloGlu. “Copulas with Directional Dependence Property.” Gazi University Journal of Science, vol. 24, no.3, July 2011, pp. 415-424.

[2] Rodriguez- Lallena, Jose Antonio, and Manuel Ubeda-Flores. “A New Class of Bivariate Copulas.” Statistics & Probability Letters, vol.66, no.3, Feb.2004, p.315

[3] Nelsen, Roger, and Manuel Ubeda-Flores. “Directional Dependence in Multivariate Distributions.” Annals of the Institute of Statistical Mathematics, vol.64, no.3, June 2012, pp. 677-685.

[4] von Eye, Alexander, and Richard P. Deshob. “Directional Dependence in Developmental Research.” International Journal of Behavior Development, vol. 36, no.4, July 2012, pp. 303-312.

[5] Muddapur, M. V. “A Remark on the Paper ‘A Note on Directional Dependence in Regression Setting’ By Engin A. Sungur.” *Communications in Statistics - Theory and Methods*, vol. 37, no. 3, 2008, pp. 386–387

[6] Schechtman, E., and S. Yitzhaki. “On the Proper Bounds of the Gini Correlation.” *Economics Letters* 63.2 (2002): 133–138. *Economics Letters*. Web.

[7] Tsukibayashi, Shomei. “The Joint Distribution and Moments of an Extreme of the Dependent Variable and the Concomitant of an Extreme of the Independent Varible.” *Communications in Statistics - Theory and Methods* 27.7 (1998): 1639–1651. *Communications in Statistics - Theory and Methods*. Web.

[8] He, Qinying, and H. N. Nagaraja. “Correlation Estimation Using Concomitants of Order Statistics from Bivariate Normal Samples.” *Communications in Statistics - Theory and Methods*. Vol. 38. N.p., 2009. 2003–2015. *Communications in Statistics - Theory and Methods*. Web.

[9] Barnett, V., P. J. Green, and A. Robinson. “Concomitants and Correlation Estimates.” *Biometrika* 63.2 (1976): 323–329. *Biometrika*. Web.