Graphical Methods For Detecting Dependence Using Copulas

Mario César Jaramillo Elorza
Associated Professor.
National University of Colombia, Medellín, mcjarami@unal.edu.co

Carlos Mario Lopera Gómez
Associated Professor.
National University of Colombia, Medellín, cmlopera@unal.edu.co

Julieth Verónica Guarín Escudero*
Student of M.Sc. in Statistics
National University of Colombia, Medellín, jvguarine@unal.edu.co

Abstract

Copulas have become a useful tool for modeling data when the dependence among random variables exists and the multivariate normality assumption is not fulfilled. The copulas have been applied in several fields. In finance, copulas are used in asset modeling and risk management. In biomedical studies, copulas are used to model correlated lifetimes and competitive risks (Escalera, G., & Carriere, J., 2003). In engineering, copulas are used in multivariate process control and hydrological modeling (Genest, C., & Favre, A., 2007). The interest in modeling multivariate problems involving dependent variables is generalized in several areas, making this methodology in a convenient way to model the dependence structure of random variables.
However, in practice there is not a standard method for selecting a copula among several possible models, so that the choice of an appropriate copula is one of the greatest challenges facing the researcher. In this paper some graphical methods for detecting dependencies among random variables are discussed.
Keywords: Copulas, Graphics, Dependence.

1. Introduction

In the theory of probability the functions called copulas actually are distribution functions representing the dependency relationships among random variables. The two-dimensional copula is a function linking two univariate distributions to construct a joint distribution function.
The copula is a convenient way to model the dependence structure of random variables (Escalera, G., & Carriere, J., 2003). This concept allows building models beyond the standards in the analysis of dependence between variables. Copula methodology captures nonlinear relationships, in particular it allows to relate extreme events occurring in nature (Nelsen, R., 2006).
The choice of the best copula within a family that captures and fits well the dependence structure is one of the problems to deal with when working with copulas. There are graphical and analytical tests to measure the fit of copula. Among the graphic tests are the Chi-Plot and the K-Plot, the former was initially proposed by Fisher, N. I., & Switzer, P. (1985) and better illustrated in Fisher, N. I., & Switzer, P. (2001). The K-Plot (Kendall Plot) was proposed by Genest, C., & Boies, J. C. (2003). Both methods are a useful tool to study the dependence between two random variables. In this work both the Chi-Plot and the K-Plot are analyzed.

2. Graphical Method for Detecting Dependence: Chi-Plot

The graph Chi-Plot was originally proposed by Fisher and Switzer (Genest, C., & Favre, A., 2007). Its construction is based on the Chi-square statistic for independence. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a random sample of a joint and continuous distribution function, H, from the random variables (X, Y), and let $I(A)$ be the characteristic function of the event A. For each observation (x_i, y_i) the following procedure is performed Moreno, D., & Blanco, L. (2012):

$$H_i = \frac{1}{n-1} \sum_{j \neq i} I(X_j \leq X_i, Y_j \leq Y_i)$$

$$F_i = \frac{1}{n-1} \sum_{j \neq i} I(X_j \leq X_i)$$

and

$$G_i = \frac{1}{n-1} I(Y_j \leq Y_i)$$

None of these quantities exclusively depend of the observations ranks. Fisher, N. I., & Switzer, P. (1985) proposed to plot the pairs (λ_i, χ_i), where:

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i (1 - F_i) G_i (1 - G_i)}}$$

and

$$\lambda_i = 4 \, \text{sign}\left(\tilde{F}_i \tilde{G}_i\right) \max\left(\tilde{F}_i^2, \tilde{G}_i^2\right)$$

Where $\tilde{F}_i = F_i - 1/2$, $\tilde{G}_i = G_i - 1/2$ for $i \in \{1, \ldots, n\}$
The Chi-Plot is a scatterplot of the pairs $(\lambda_i, \chi_i), i = 1, \ldots, n$ where λ_i is a measure of distance from the observation (X_i, Y_i) to data center (Moreno, D., & Blanco, L., 2012).
According to Del Rio, R., & Quesada, V. (2007) all values of λ_i must be in the interval $[-1; 1]$. If the data constitute a bivariate sample with independent continuous marginals, the values of λ_i will be evenly distributed. However, if X and Y are associated, the values of λ_i will show forming groups, in particular, positive values of λ_i indicate that X_i and Y_i are relatively larger or smaller (at the same time) than the median, while λ_i negative correspond to X_i and Y_i located on opposite sides relative to their median (Moreno, D., & Blanco, L., 2012).

3. Graphical Method for Detecting Dependence: K-Plot

The K-plot (Kendall-plot) was created by Genest, C., & Boies, J. C. (2003). This tool also is based on the observations ranks using the integral transformation of multivariate probabilities, producing a similar graph to QQ-plot conventional (Moreno, D., & Blanco, L., 2012).
Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a random sample of a joint and continuous distribution function, H, from the random variables (X, Y). To build the K-Plot must proceed as follows:

1. For each $1 \leq i \leq n$ compute H_i (as in the Chi-Plot).
2. Sort the H_i values such that $H_{(1)} \leq \ldots \leq H_{(n)}$.
3. Plot the pairs $(W_{i:n}, H_i)$, where $W_{i:n}$ is the expectation of the ith order statistic in a sample of size n, which is calculated as follows:

$$W_{i:n} = n \binom{n-1}{i-1} \int_0^1 w \left[K_0(w)\right]^{i-1} \left[1 - K_0(w)\right]^{n-i} dK_0(w)$$

with

$$K_0(w) = w - w \log(w) \quad 0 \leq w \leq 1$$

When the scatter plot of $H_{(i)}$ against $W_{i:n}$ moves away from the diagonal, then one can to assume that there is a functional dependence between the two variables involved.

4. Simulation

In this section we present a simulation of variables with a high or low level of dependence for different sample sizes. In addition, we show the implementation of Chi-plot and K-plot through the package CDVine of R.

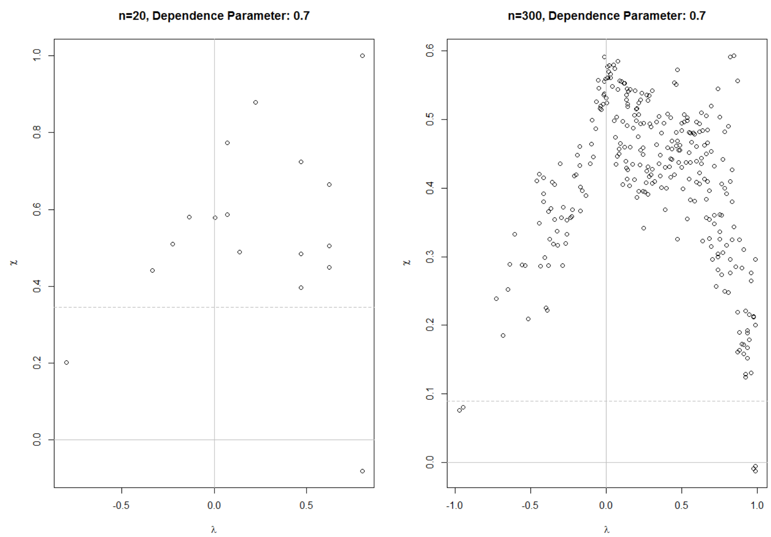

Figure 1: Chi-Plot with sample sizes n=20,300. τ =0.7

Figure 1. shows simulated data of a Gaussian copula with dependence parameter equal to 0.7. When the sample size is small, the Chi-Plot shows no clear evidence that data are dependent. On the other hand, the K-Plot (Figure 2) visualizes the dependence when there are few data.

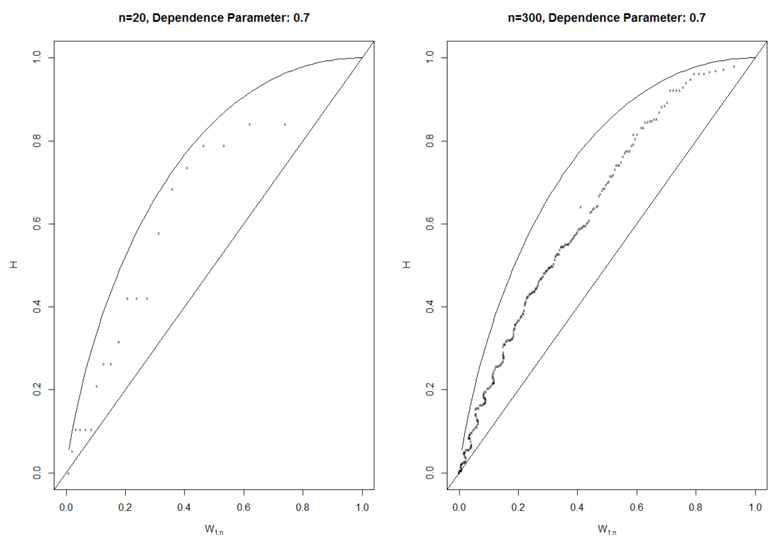

Figure 2: K-Plot with sample sizes n=20,300. τ =0.7

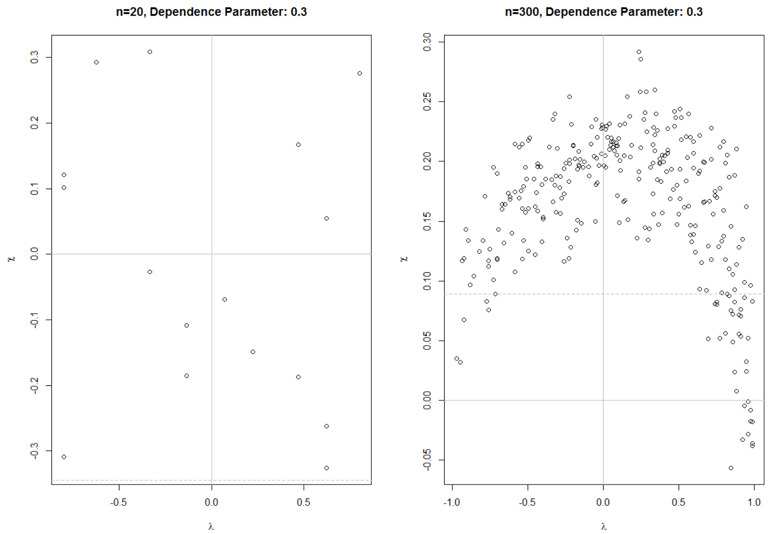

Figure 3: Chi-Plot with sample sizes n=20,300. τ =0.3

When variables have low dependence, Chi-Plot (Figure 3) does not show a clear trend, while the K-Plot (Figure 4) shows clearly a low dependence on the data, since data fit to a straight line.

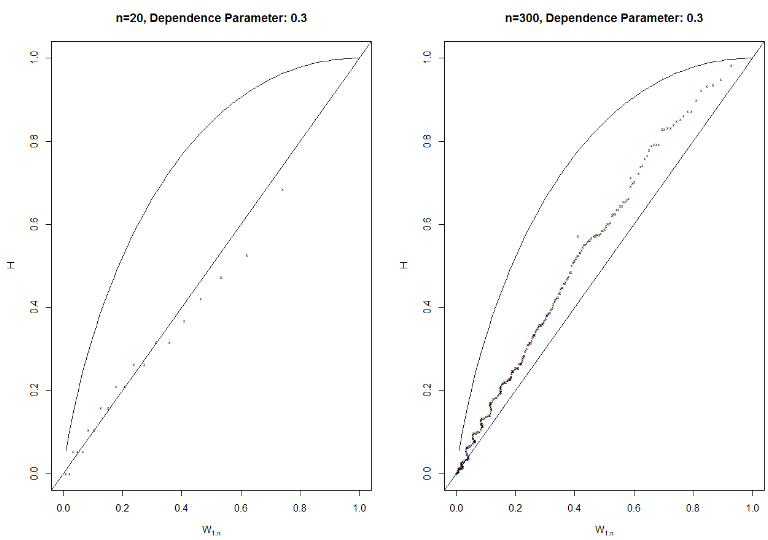

Figure 4: K-Plot with sample sizes n=20,300. τ =0.3

5. Conclusions

- Although there are analytical tests to prove dependency, graphical methods provide a visual and descriptive data that helps people who do not work in the area of statistics.
- The K-Plot shows clearly the low or high dependence between variables than the Chi-Plot when there are small sample sizes.

References

Escarela, G., & Carriere, J. F. (2003). Fitting competing risks with an assumed copula. Statistical Methods in Medical Research, vol. 12, pp. 333-349.

Genest, C. & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, pp. 347-368.

Nelsen, R. (2006). An Introduction to Copulas. Springer Series in Statistics.

Fisher, N. I., & Switzer, P. (1985). Chi-plots for assessing dependence. Biometrika, vol. 72, no. 2, pp. 253-265.

Fisher, N. I., & Switzer, P. (2001). Graphical assessment of dependence: Is a picture worth 100 tests?. The American Statistician, vol. 55, no. 3, pp. 233-239.

Genest, C., & Boies, J. C. (2003). Detecting dependence with kendall plots. The American Statistician, vol. 57, no. 4, pp. 275-284.

Moreno, D., & Blanco, L. (2012). Método para elegir una Cópula Arquimediana Óptima. Master's thesis, Universidad Nacional de Colombia.

Del Rio, R., & Quesada, V. (2007). Teoría de Cópulas y control de Riesgo Financiero. PhD thesis, Universidad Complutense de Madrid.