Avestia Publishing
Journal of Machine Intelligence and Data Science (JMIDS)
Volume 4, Year 2023
ISSN: 2564-3282
DOI: 10.11159/jmids.2023.001

Practical Representations of Copula and Joint Density Estimates

Yishan Zang[1], Serge Provost[1]
[1]The University of Western Ontario, Department of Statistical and Actuarial Sciences
1151 Richmond Street, London, Ontario, Canada N6A 3K7
yzang8@uwo.ca; provost@uwo.ca

***Abstract** - A moment-based approximation methodology for estimating a copula density from bivariate observations is introduced. The resulting simple representation of the copula density is suitable for reporting purpose or carrying out further algebraic manipulation. Empirical copula density functions will also be determined from kernel density estimates. A technique for obtaining a joint density from marginal density estimates and a copula density is proposed as well. The Bernstein copula density approximants will be utilized for comparison purposes. The results are applied to two stocks' closing prices and a stock's price and its running maximum. In the latter case, the model is related to a Brownian motion process.*

Keywords: Empirical copulas, Bivariate density estimation, Data modelling, Brownian motion process, Financial application.

1. Introduction

Copulas are principally utilized for modelling dependencies in multivariate distributions. Their properties have been increasingly exploited in numerous types of scientific investigations; for instance, the reader may refer to Chao et al. [1], Carrera et al. [2] and Chen et al. [3] for recent advances in the area of artificial intelligence. The key idea behind copulas is that the joint distribution of two or more variables can be represented in terms of their marginal distributions and a specific correlation structure. As a measure of dependence, they have for instance found applications in reliability theory, signal processing, geodesy, hydrology and medicine. Results involving empirical bivariate copula densities are discussed in this paper.

Let $F(x_1, x_2)$ be the joint cumulative distribution function of random variables X_1 and X_2 having continuous marginal distribution functions $F_1(x_1)$ and $F_2(x_2)$. According to Sklar [4], there exists a unique bivariate copula $C: I^2 \mapsto I$ (the unit interval) such that

$$F(x_1,\, x_2) = C\big(F_1(x_1),\, F_2(x_2)\big), \tag{1}$$

where $C(\cdot,\cdot)$ is a joint cumulative distribution function having uniform marginals. Conversely, for any continuous cumulative distribution function $F_1(x_1)$ and $F_2(x_2)$ and any copula C, the function F defined in Equation (1) is a joint distribution function with marginal distributions F_1 and F_2.

Sklar's theorem provides a technique for constructing copulas. Indeed, the function

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1),\, F_2^{-1}(u_2)\right) \tag{2}$$

is a bivariate copula, where the quasi-inverse F_i^{-1} for $i = 1, 2$ is defined by

$$F_i^{-1}(u) = \inf\{x | F_i(x) \geq u\} \;\; \forall u \in (0,1). \tag{3}$$

We shall denote the probability density function corresponding to the copula $C(u_1, u_2)$ by

Date Received: 2022-06-29
Date Revised: 2023-05-18
Date Accepted: 2023-06-24
Date Published: 2023-07-05

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2). \quad (4)$$

The following relationship between the joint density function $f(\cdot,\cdot)$ and the copula density function $c(\cdot,\cdot)$ can readily be obtained from Equation (1):

$$f(x_1, x_2) = f_1(x_1) f_2(x_2)\, c\big(F_1(x_1), F_2(x_2)\big), \quad (5)$$

where $f_1(x_1)$ and $f_2(x_2)$ respectively denote the marginal density functions of X_1 and X_2. Accordingly, the copula density function can be expressed as follows:

$$c(u_1, u_2) = \frac{f\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right)}{f_1\left(F_1^{-1}(u_1)\right) f_2\left(F_2^{-1}(u_2)\right)}. \quad (6)$$

The proposed approaches to estimating copula density functions are summarized in Sections 2 and 3. The estimation of a joint density function in terms of marginal distributions and a copula density estimate is discussed in Section 4. Three illustrative examples are provided in the last section.

2. Copula Density Based on Kernel Density Estimation

On applying the kernel density estimation (kde) method to a given bivariate dataset, one can obtain an estimate of the bivariate probability density of $\boldsymbol{X}$. In light of Equation (6), the copula density can be represented as follows:

$$f_c(u_1, u_2) = \frac{f_{\mathbf{X}}\left(Q_{X_1}(u_1), Q_{X_2}(u_2)\right)}{f_{X_1}\left(Q_{X_1}(u_1)\right) f_{X_2}\left(Q_{X_2}(u_2)\right)}, \quad (7)$$

where $f_{\mathbf{X}}$ can be estimated by a bivariate kde and $Q_{X_i}(\cdot)$ denotes the quantile function.

The marginal density function of each variable can be obtained by determining their respective kernel density estimates. The inverse cdfs Q_{X_1} and Q_{X_2} can be determined in polynomial form by making use of a moment-based method or the method of least squares.

3. A Moment-based Bivariate Polynomial Approximation of the Copula Density

Once a copula density is determined from Equation (7), it can be approximated by the product of a base density and a bivariate polynomial whose coefficients are obtained from the joint moments associated with the copula density.

The proposed procedure for achieving this is described in the next result.

Result 1 *Let $f_X(x_1, x_2)$ be the density function of a bivariate continuous random vector $\boldsymbol{X} = (X_1, X_2)$ defined in the rectangle $(l_1, u_1) \times (l_2, u_2)$, whose joint moments are*

$$\mu_{\boldsymbol{X}}(i, j) \equiv \int_{l_1}^{u_1} \int_{l_2}^{u_2} x_1^i \; x_2^j \, f_X(x_1, x_2) dx_2 dx_1. \quad (8)$$

Let $\psi(x_1, x_2)$ be an initial joint density estimate, whose joint moments are

$$m_X(i, j) \equiv \int_{l_1}^{u_1} \int_{l_2}^{u_2} x_1^i \, x_2^j \;\; \psi(x_1, x_2) dx_2 dx_1. \quad (9)$$

Assuming that the sequence $\mu_X(i, j)$, $i = 0,1,2, \dots$, $j =$ 0,1,2, uniquely defines the distribution of $\boldsymbol{X}$, the density function of $\boldsymbol{X}$ can be approximated by

$$f_n(x_1, x_2) = \psi(x_1, x_2) \sum_{i=0}^{n} \sum_{j=0}^{n} \xi_{i,j} \, x_1^i x_2^j, \quad (10)$$

where $\xi_{i,j}$ can be determined by letting

$$\int_{l_1}^{u_1} \int_{l_2}^{u_2} y_1^h y_2^g \, f_{\boldsymbol{X}}(x_1, x_2) dx_2 dx_1 = \int_{l_1}^{u_1} \int_{l_2}^{u_2} \psi(x_1, x_2) \sum_{i=0}^{n} \sum_{j=0}^{n} \xi_{i,j} \, x_1^{i+h} x_2^{j+g} \, d\, x_2 d\, x_1, \quad (11)$$

$h = 0, 1, \dots, n;\ g = 0, 1, \dots, n$, or equivalently,

$$\mu_X(h, g) = \sum_{i=0}^{n} \sum_{j=0}^{n} \xi_{i,j} m_X(i + h, j + g), \quad (12)$$

$h = 0, 1, \dots, n;\ g = 0, 1, \dots, n$. Thus, we can obtain the polynomial coefficients $\xi_{i,j}$ and $f_n(x_1, x_2)$ from the moments of $f_X(\cdot)$ and $\psi(\cdot)$ by solving the linear system specified by Equation (12).

The degree n used in the polynomial adjustment should be selected so that f_n provides an accurate approximation to the estimate of the copula density, which can be determined for instance by evaluating their integrated squared differences.

4. Estimating a Joint Density from the Marginal Density Estimates and the Copula Density

On applying Equation (5), that is,

$$f(x_1, x_2) = f_1(x_1) f_2(x_2)\, c\big(F_1(x_1), F_2(x_2)\big), \quad (5)$$

one can determine the joint density where for instance, $c(\cdot,\cdot)$ could be taken as a Bernstein copula density which is described in Sancetta and Satchell [5].

Suppose that observations are available on the random variables X_1 and X_2. The marginal densities $f_1(x_1)$ and $f_2(x_2)$ are estimated and a Bernstein's copula density of high order is determined. A joint density estimate of (X_1, X_2) can then be obtained via Equation (5).

5. Applications

The results are applied to two stocks' closing prices as well as a stock's price and its running maximum.

5.1. Copula Density Estimation Methodologies Applied to Two Stocks

The two stocks selected are GOOG (Alphabet Inc.) and AAPL (Apple Inc.). The bivariate data are the daily closing prices of (GOOG, AAPL) in 2019. Each component of the data has been standardized. The joint KDE of $\boldsymbol{X}$, the marginal densities of X_1 and X_2 and the corresponding cdf inverses are displayed in Figures 1-5.

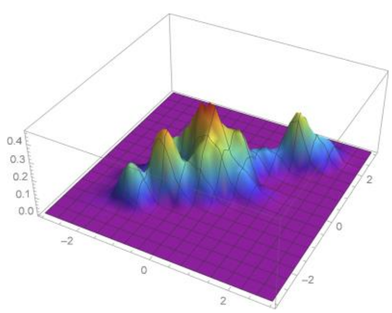

Figure 1. Bivariate kde of X.

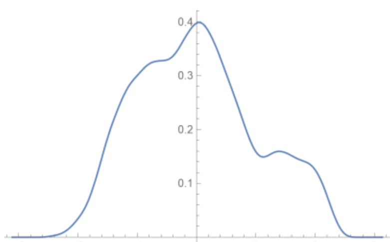

Figure 2. Marginal density of X_1.

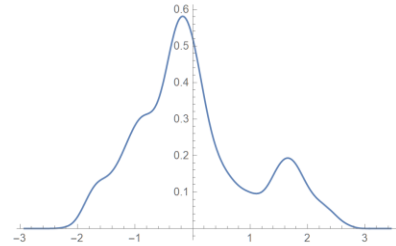

Figure 3. Marginal density of X_2.

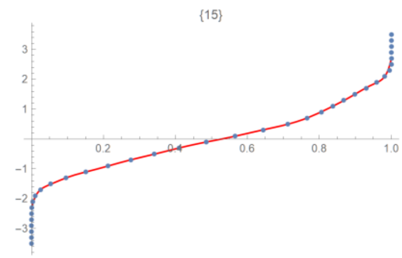

Figure 4. Estimate of inverse cdf Q_{X_1}.

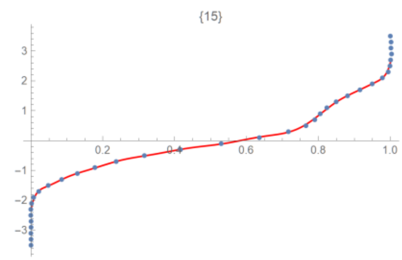

Figure 5. Estimate of inverse cdf Q_{X_2}.

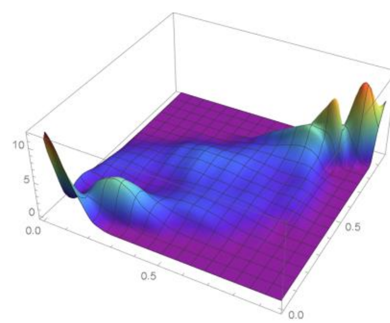

Figure 6. KDE based copula density.

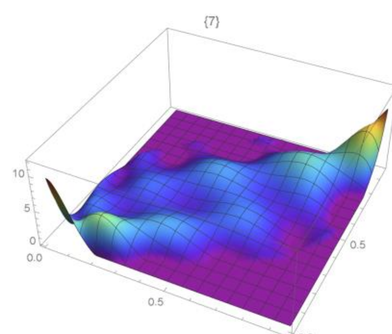

Figure 7. Moment-based bivariate polynomial estimate of the copula density.

The resulting KDE based (Section 2) and moment-based (Section 3) copula densities plotted in Figures 6 and 7 are seen to be similar.

5.2. Determination of a Joint Density Estimate from the Marginal Density Estimates and the Copula Density

Using the two stocks' standardized data, a Bernstein's copula density of degree 250, which is shown in Figure 8, is utilized as an estimate of the underlying copula density. The density estimate (Figure 9) resulting from the application of the methodology described in Section 4 and a KDE (Figure 10) exhibit very similar features.

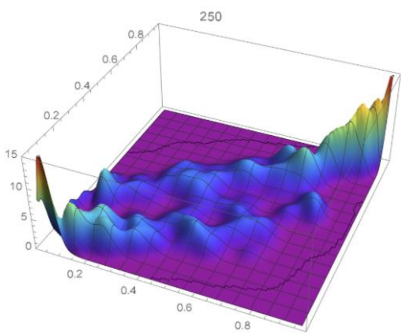

Figure 8. Bernstein's copula density with degree 250.

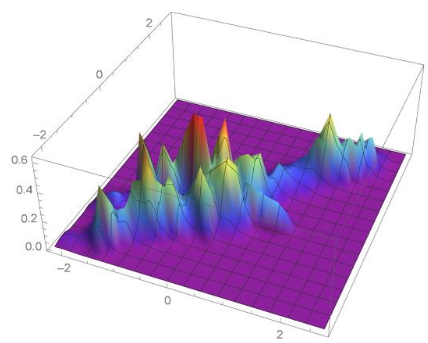

Figure 9. The estimated joint density.

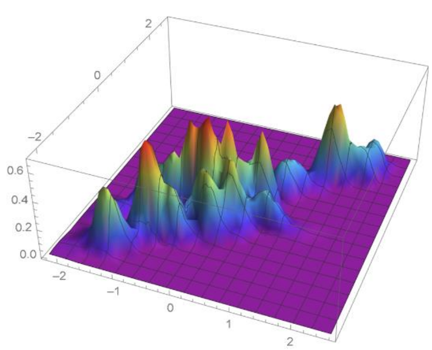

Figure 10. Bivariate KDE of X.

5.3. Copula Associated with a Brownian Motion Process and Its Running Maximum

The data consists of the daily closing prices of AC.TO (Air Canada) during 2019. To relate the data to a standard Wiener process, the first data point should be zero, the differences between successive observations should ideally often change signs and have a variance of one, and there should be one unit of time between successive observations. Hence the following transformation is used.

Let $U_1, U_2, \ldots, U_n$ denote the closing prices and $V_1, V_2, \ldots, V_{n-1}$ be the differences between successive closing prices, that is, $V_i = U_{i+1} - U_i$; denoting by σ_D the standard deviation of the differences $V_1, V_2, \ldots, V_{n-1}$, the following transformation is applied $W_i = \frac{U_i - U_1}{\sigma_D}$ and the resulting data is denoted by $W_1, W_2, \ldots, W_n$.

Let Z_i be the i^{th} running maximum, that is, $Z_i = \text{Max}\{W_1, W_2, \ldots, W_i\}$, $i = 1, 2, \ldots, n$. Then, the resulting bivariate data, (W_i, Z_i), $i = 1, 2, \ldots, n$, has the features of a Brownian motion process and its running maximum.

The W_i's and Z_i's are plotted in Figures 11 and 12. The resulting copula density as obtained by applying the methodology described in Section 2, is shown in Figure 13.

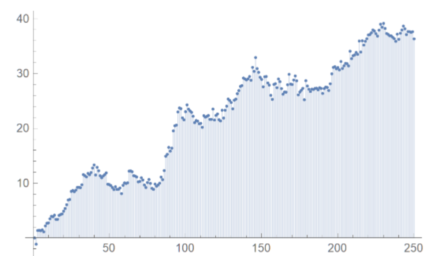

Figure 11. List Plot of the W_i's.

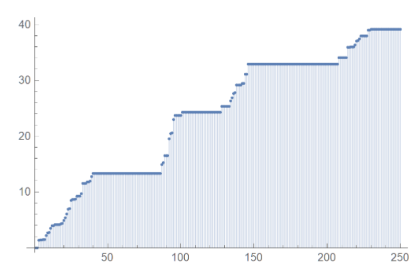

Figure 12. List Plot of the Z_i's.

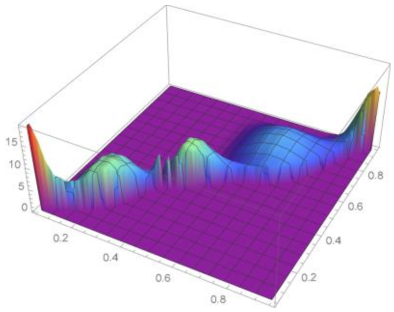

Figure 13. The estimated copula density.

6. Conclusion

As copula density estimates are usually expressed in complicated forms, the bivariate polynomial approximation that is proposed in this paper ought to prove more suitable for reporting purposes. Approximations by means of Bernstein polynomials and the kernel density estimation approach are discussed as well. Additionally, a flexible technique for estimating joint density functions is introduced. The proposed methodologies were successfully applied to two stocks' closing prices as well as a set of observations and its running maximum.

References

[1] M. Chao, S. Z. Xin, and L. S. Min, "Neural network ensembles based on copula methods and distributed multiobjective central force optimization algorithm," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 203–212, 2014.

[2] D. Carrera, R. Santana, and J. A. Lozano, "Vine copula classifiers for the mind reading problem," *Progress in Artificial Intelligence*, vol. 5(4), pp. 289–305, 2016.

[3] T. Chen, S.-S. He, J.-Q. Wang, L. Li, and H. Luo, "Novel operations for linguistic neutrosophic sets on the basis of Archimedean copulas and co-copulas and their application in multi-criteria decision-making problems," *Journal of Intelligent & Fuzzy Systems*, vol. 37(2), pp. 2887–2912, 2019.

[4] M. Sklar, "Fonctions de répartition à *n* dimensions et leurs marges," *Publications de I'Institut de statistique de I'Université de Paris*, vol. 8, pp. 229-231, 1959.

[5] A. Sancetta and S. Satchell, "The Bernstein copula and its applications to modelling and approximations of multivariate distributions," *Econometric Theory*, vol. 20, pp. 535-562, 2004.