

Research Article

Why Are FGM Copulas Successful? A Simple Explanation

Songsak Sriboonchitta¹ and Vladik Kreinovich ²

¹Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

²Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA

Correspondence should be addressed to Vladik Kreinovich; vladik@utep.edu

Received 17 March 2017; Accepted 22 March 2018; Published 14 May 2018

Academic Editor: Katsuhiko Honda

Copyright © 2018 Songsak Sriboonchitta and Vladik Kreinovich. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the most computationally convenient nonredundant ways to describe the dependence between two variables is by describing the corresponding copula. In many applications, a special class of copulas—known as FGM copulas—turned out to be most successful in describing the dependence between quantities. The main result of this paper is that these copulas are the fastest to compute, and this explains their empirical success. As an auxiliary result, we also show that a similar explanation can be given in terms of fuzzy logic.

1. Introduction

What Is a Copula? A Brief Reminder. In many practical situations, we know the distribution of each of the two random variables X and Y , and we now need to also describe their joint distribution.

The distribution of each of the random variables can be described by the corresponding cumulative distribution functions $F_X(x) \stackrel{\text{def}}{=} \text{Prob}(X \leq x)$ and $F_Y(y) \stackrel{\text{def}}{=} \text{Prob}(Y \leq y)$.

Similarly, to describe their joint distribution, we can use corresponding 2D cumulative distribution function (cdf)

$$F_{XY}(x, y) \stackrel{\text{def}}{=} \text{Prob}(X \leq x \wedge Y \leq y). \quad (1)$$

In principle, we can thus try to determine the values $F_{XY}(x, y)$ corresponding to all possible pairs (x, y) . However, from the practical viewpoint, this is redundant; indeed

- (i) the 2D cdf $F_{XY}(x, y)$ also contains information about the 1D cdfs $F_X(x)$ and $F_Y(y)$, as $F_X(x) = F_{XY}(x, +\infty)$ and $F_Y(y) = F_{XY}(+\infty, y)$,
- (ii) so if we determine all the values $F_{XY}(x, y)$, we will also be determining the values $F_X(x)$ and $F_Y(y)$, but
- (iii) we consider the cases when the 1D cdf values are already known, so soliciting them again is unnecessary.

It is therefore desirable to describe the dependence between X and Y in a nonredundant way, so that

- (i) from this description, we will not be able to extract the known 1D cdfs, but
- (ii) from this information and from the 1D cdfs, we will be able to extract the 2D cdf.

Such a nonredundant description is indeed known, it is a *copula* $C(u, v)$, a function from $[0, 1] \times [0, 1]$ to $[0, 1]$, for which, for all real numbers x and y , we have

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)); \quad (2)$$

see, for example, [1–5].

Properties of Copulas. Not every function $C(u, v)$ is a copula for an appropriate 2D distribution. For a function to be a copula, it has to satisfy some properties. In this paper, we will use the following properties, which can be easily derived from the definition of the copula:

$$\begin{aligned} C(0, v) &= C(u, 0) = 0; \\ C(1, v) &= v; \\ C(u, 1) &= u. \end{aligned} \quad (3)$$

FGM Copulas and Their Success. There exist many different copulas. Interestingly, in many practical applications, the following Farlie–Gumbel–Morgenstern (FGM) copula turns out to be very successful:

$$C(u, v) = u \cdot v + \theta \cdot u \cdot (1 - u) \cdot v \cdot (1 - v) \quad (4)$$

for $\theta \in [-1, 1]$. The original papers are [6–8]; see, for example, [9, 10] and references therein for latest results.

Why? To the best of our knowledge, until now, there was no convincing explanation of why FGM copulas are so empirically successful. In this paper, we provide such an explanation.

2. Materials and Methods

2.1. Explanation Based on Computational Complexity:

Main Result

Statistical Data Processing Is Computing. Statistical data processing involves a large amount of computing. With the ever increasing amount of data, processing all this data requires more and more computation time—often to the extent that we exceed the capabilities of our computers.

From this viewpoint, it is desirable to select techniques which are as computationally efficient as possible. With respect to copulas, this means that we should select copulas $C(u, v)$ whose values are the easiest (thus, the fastest) to compute.

Which Functions Are the Fastest to Compute? In the computers, the only exactly hardware supported operations are addition, subtraction, and multiplication. Everything else—from division to special functions such as $\exp(x)$, $\sin(x)$, and so on—is approximated by a sequence of elementary hardware supported operations. The more accuracy we need, the more elementary operations we need, and, thus, the longer the corresponding computations will be.

Therefore, the fastest-to-compute functions are functions that can be exactly represented as a sequence of elementary operations: in this case, the number of elementary operations remains the same no matter what accuracy we desire in our computations. In other words, we are looking for functions which can be obtained from constants and original quantities x_1, \dots, x_n by applying addition, subtraction, and multiplication. One can easily see that such functions are *polynomials*; indeed

- (i) every polynomial is a sum of monomials, and each monomial is a product of a constant and variables, so each polynomial is indeed a superposition of additions and multiplications;
- (ii) and vice versa, each constant and each variable are polynomials, and the sum, the difference, and the product of two polynomials are also polynomials; thus, by induction, we can prove that every superposition of addition, subtraction, and multiplication is a polynomial.

Not all polynomials are equally easy or equally difficult to compute. Out of the three elementary operations, the most time-consuming operation is multiplication. Thus, the fewer the multiplications are, the faster the computation of the corresponding function is.

- (i) With one multiplication—performed in parallel—we can compute linear functions $a_0 + \sum_{i=1}^n c_i \cdot x_i$ and also products $x_i \cdot x_j$ of two variables.
- (ii) By applying second multiplication to the results of the first one, we can thus compute 3rd degree polynomials—or products of 4 variables and so on.

In general, the higher the degree is, the more the time is needed to compute the corresponding polynomial.

Resulting Idea. From the viewpoint of selecting fastest-to-compute copulas, we should select polynomial copulas and, among them, copulas of the smallest possible degree.

Let us describe the results of such a selection.

Proposition 1. *Every polynomial copula has the form*

$$C(u, v) = u \cdot v + \theta(u, v) \cdot u \cdot (1 - u) \cdot v \cdot (1 - v), \quad (5)$$

for some polynomial $\theta(u, v)$.

Comments

- (i) For reader's convenience, the proof is placed in the special proof section.
- (ii) As a consequence of this proposition, we get the following results.

Corollary 2. *The only polynomial copula of 3rd degree is $C(u, v) = u \cdot v$.*

Comment. This copula is actually of 2nd degree; it corresponds to the case of two independent variables. Thus, to describe dependence, we need to consider polynomials of higher degree.

Corollary 3. *The only polynomial copulas of 4th degree are FGM copulas.*

Comments

- (i) This result explains the empirical success of the FGM copulas: among copulas describing true dependence, they are the easiest to compute.
- (ii) Since the FGM copulas are symmetric $C(u, v) = C(v, u)$, asymmetric dependence requires higher-degree polynomial copulas.
- (iii) An alternative explanation of the FGM formulas, based on *fuzzy logic*, is given in the next subsection.

2.2. *Explanation Based on Computational Complexity: Proof of the Main Result.* (1) The first condition on the copula, the

condition that $C(0, \nu) = 0$ for all ν , means that if $u = 0$, then $C(u, \nu) = 0$.

An arbitrary polynomial $C(u, \nu)$ can be represented as

$$C(u, \nu) = C_0(\nu) + u \cdot C_1(u, \nu), \quad (6)$$

where $C_0(\nu)$ is the sum of all the monomials that do not contain u and $C_1(u, \nu)$ is the result of dividing all u -containing monomials by u .

For $u = 0$, the condition $C(0, \nu) = 0$ means that $C_0(\nu) = 0$ for all ν . Thus, $C(u, \nu) = u \cdot C_1(u, \nu)$ for some polynomial $C_1(u, \nu)$.

(2) The condition $C(u, 0) = u \cdot C_1(u, 0) = 0$ for all $u \neq 0$ implies that $C_1(u, 0) = 0$ for all u and, thus, that $C_1(u, \nu) = \nu \cdot C_2(u, \nu)$ for some function $C_2(u, \nu)$. Therefore,

$$C(u, \nu) = u \cdot C_1(u, \nu) = u \cdot \nu \cdot C_2(u, \nu). \quad (7)$$

(3) The condition $C(1, \nu) = \nu$ takes the form $\nu \cdot C_2(1, \nu) = \nu$, so $C_2(1, \nu) = 1$, and so $f(u, \nu) \stackrel{\text{def}}{=} C_2(u, \nu) - 1 = 0$ when $u = 1$, that is, when $1 - u = 0$.

(4) Similarly to part 1 of this proof, this implies that

$$C_2(u, \nu) - 1 = (1 - u) \cdot C_3(u, \nu) \quad (8)$$

for some polynomial $C_3(u, \nu)$. Similarly, the condition $C(u, 1) = 1$ implies that $C_3(u, \nu) = (1 - \nu) \cdot C_4(u, \nu)$ for some polynomial $C_4(u, \nu)$. Thus,

$$\begin{aligned} C_2(u, \nu) - 1 &= (1 - u) \cdot C_3(u, \nu) \\ &= (1 - u) \cdot (1 - \nu) \cdot C_4(u, \nu), \end{aligned} \quad (9)$$

and hence

$$\begin{aligned} C_2(u, \nu) &= 1 + (1 - u) \cdot (1 - \nu) \cdot C_4(u, \nu), \\ C(u, \nu) &= u \cdot \nu \cdot C_2(u, \nu) \\ &= u \cdot \nu \cdot (1 + (1 - u) \cdot (1 - \nu) \cdot C_4(u, \nu)). \end{aligned} \quad (10)$$

This is the desired formula, with $\theta(u, \nu) = C_4(u, \nu)$.

The proposition is proven.

2.3. Explanation Based on Fuzzy Logic

What Is Fuzzy Logic? A Brief Reminder. An alternative explanation comes from *fuzzy logic*, where numbers from the interval $[0, 1]$ describe the expert's degree of confidence in a statement. Fuzzy logic was invented by Zadeh [11]; for the state-of-the-art, see, for example, [12–15].

In fuzzy logic, once we know the expert's degree of confidence a in a statement A , his/her degree of confidence in its negation $\neg A$ is estimated as $1 - a$.

Similarly, if we know the expert's degree of confidence a in a statement A and we know the expert's degree of confidence b in a statement B , then the expert's degree of confidence in a conjunction $A \wedge B$ is estimated as $f_{\wedge}(a, b)$ for an appropriate function $f_{\wedge}(a, b)$; this function is known as an “and”-operation or a *t-norm*. One of the most widely used “and”-operations is the algebraic product $f_{\wedge}(a, b) = a \cdot b$ – that

corresponds to the situation when A and B are statistically independent and we take probability as degree of confidence. This is the “and”-operation that we will use in this section.

Similarly, to estimate the expert's degree of confidence in a statement $A \vee B$, we apply an appropriate “or”-operation $f_{\vee}(a, b)$ (also called *t-conorm*) to the corresponding degrees a and b . One of the most widely used “or”-operations is $f_{\vee}(a, b) = \min(a + b, 1)$. This is the “or”-operation that we will use in this section.

Copula as a Particular Case of an “and”-Operation. A copula can also be viewed as an “and”-operation: it transforms the probabilities $F_X(x) = \text{Prob}(X \leq x)$ and $F_Y(y) = \text{Prob}(Y \leq y)$ of the events $X \leq x$ and $Y \leq y$ into the probability $F_{XY}(x, y) = \text{Prob}(X \leq x \wedge Y \leq y)$ that the first event occurs *and* the second event occurs. How can we go from the original “crisp” “and”-operation to a new “fuzzy” one?

Towards a Fuzzy Explanation of the FGM Copula. For each of the two statements A and B , we want to cover both possibilities:

- (i) that the corresponding statement is absolutely true,
- (ii) that the corresponding statement is “fuzzy,” that is, to some extent true and to some extent false.

In other words, fuzzy means that there is some degree of belief that A is true *and* that its negation is true.

Thus, we can say that the statement $A \wedge B$ is true if

- (i) either A and B are absolutely true,
- (ii) or A and B are both “fuzzy,” that is, true to some extent and false to some extent.

The degree to which A is true is a . Thus, the degree to which the negation $\neg A$ is true is $1 - a$. Therefore, the degree to which both the statement A and its negation are both true is $a \cdot (1 - a)$. This is a degree to which the statement A is fuzzy.

Similarly, the degree to which B is fuzzy is equal to $b \cdot (1 - b)$. Thus, the degree to which both A and B are fuzzy is equal to the product $a \cdot (1 - a) \cdot b \cdot (1 - b)$.

If we denote the degree to which this both-fuzzy case contributes to “and” by θ , then the contribution of this case to the overall trueness of the conjunction $A \wedge B$ is $\theta \cdot a \cdot (1 - a) \cdot b \cdot (1 - b)$.

The degree to which both A and B are true can be estimated as $a \cdot b$. Thus, if we use $\min(a + b, 1)$ as the “or”-operation, then the resulting overall degree has the desired form

$$a \cdot b + \theta \cdot a \cdot (1 - a) \cdot b \cdot (1 - b). \quad (11)$$

(at least while this sum does not exceed 1, and for the FGM copulas, it does not exceed 1.)

Therefore, we indeed have an alternative—fuzzy-logic-based—explanation of the FGM copula.

Comment. For general aspects of relation between fuzzy and copulas, see, for example, [16–18].

3. Discussion and Conclusion

Problem: Reminder. In many practical applications, correlation is used to describe dependence between random variables. However, correlation only captures possible linear dependence between random variables. To describe a general—possibly nonlinear—dependence, we need to use, for example, the copula techniques.

There exist many different families of copulas. It turns out that, in many applications, the actual dependence between random variables is best described by copulas from a special family of FGM copulas. Up to now, there have been no convincing explanations for this empirical observation.

Our Results. In this paper, we provide two possible theoretical explanations for this empirical phenomenon. First, we show that the FGM copulas are the easiest to compute—this is one possible explanation for their empirical success. Second, we show that these copulas naturally appear when we use fuzzy logic to formalize our imprecise understanding of how to describe the dependence between random variables.

Discussion. The fact that these two explanations lead to the same class of empirically successful copulas makes us confident that this is indeed the best possible class.

Our results will also, hopefully, make practitioners and researchers more confident that FGM copulas are indeed the best and, thus, encourage them to use these copulas even more.

Remaining Open Problems. An interesting open problem is related to the fact that the FGM family of copulas is a 1-parametric family. This family may be the most accurate approximator among all 1-parametric families, but the general dependence can be more complex than this. Thus, to get an even more accurate description of the dependence between several variables, it is desirable to use 2- and more-parametric families. Which 2-, 3-, . . . , parametric families should we use?

Can we use computational complexity-related ideas to come up with appropriate multidimensional families of copulas? Our arguments imply that all elements of such families should be polynomials of higher order, but what exactly are the formulas that we should use? Can we use fuzzy logic to transform our informal understanding of this problem into precise formulas for such families? Or do we need new methods for that? This would be interesting to investigate. A good start would be to first analyze this problem empirically: Which 2-parametric families of copula are empirically the best?

Disclosure

A preliminary version of this paper was posted to the University of Texas at El Paso Technical Report UTEP-CS-17-24.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

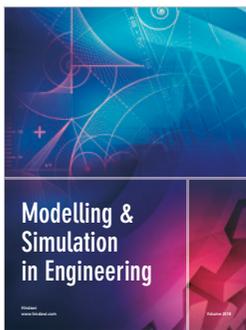
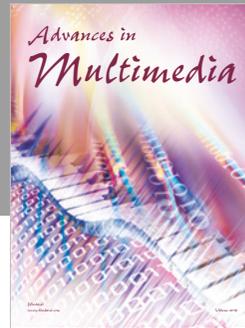
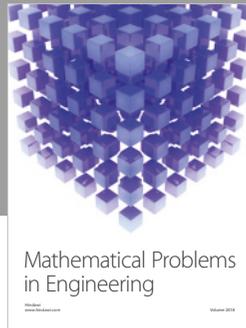
Acknowledgments

This work was supported by the Center of Excellence in Econometrics, Chiang Mai University, Thailand. It was also supported in part by the National Science Foundation Grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721 and by an award “UTEP and Prudential Actuarial Science Academy and Pipeline Initiative” from Prudential Foundation. The authors are greatly thankful to all the participants of the 2017 International Conference of the Thailand Econometric Society TES’2017, especially to Zheng Wei for valuable discussions and to the anonymous referees for important suggestions.

References

- [1] P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, Eds., *Copula Theory and Its Applications*, Springer Verlag, Berlin, Heidelberg, New York, NY, USA, 2010.
- [2] V. Kreinovich, H. T. Nguyen, S. Sriboonchitta, and O. Koshel'eva, “Why copulas have been successful in many practical applications: a theoretical explanation based on computational efficiency,” in *Integrated Uncertainty in Knowledge Modeling and Decision Making, Proceedings of The Fourth International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (IUKM'15)*, V.-N. Huynh, M. Inuiguchi, and T. Denoeux, Eds., vol. 9376 of *Springer Lecture Notes in Artificial Intelligence*, pp. 112–125, Nha Trang, Vietnam, 2015.
- [3] J.-F. Mai and M. Scherer, *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, World Scientific, Singapore, 2017.
- [4] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press, Princeton, New Jersey, USA, 2015.
- [5] R. B. Nelsen, *An Introduction to Copulas*, Springer Verlag, Berlin, Heidelberg, New York, NY, USA, 2007.
- [6] D. J. G. Farlie, “The performance of some correlation coefficients for a general bivariate distribution,” *Biometrika*, vol. 47, pp. 307–323, 1960.
- [7] E. J. Gumbel, “Bivariate exponential distributions,” *Journal of the American Statistical Association*, vol. 55, pp. 698–707, 1960.
- [8] D. Morgenstern, “Einfache beispiele zweidimensionaler verteilungen,” *Mitteilungsblatt für Mathematische Statistik*, vol. 8, pp. 234–235, 1956.
- [9] V. Kreinovich, S. Sriboonchitta, and V. N. Huynh, Eds., *Robustness in Econometrics*, Springer Verlag, Cham, Switzerland, 2017.
- [10] Z. Wei, D. Kim, T. Wang, and T. Teetranont, “A multivariate generalized FGM copulas and its application to multiple regression,” in *Robustness in Econometrics*, V. Kreinovich, S. Sriboonchitta, and V. N. Huynh, Eds., vol. 692 of *Studies in Computational Intelligence*, pp. 363–380, Springer Verlag, Cham, Switzerland, 2017.
- [11] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [12] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, NY, USA, 2017.
- [13] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, NJ, USA, 1995.

- [14] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
- [15] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman & Hall, Boca Raton, Fla, USA, 3rd edition, 2006.
- [16] P. Hájek and R. Mesiar, "On copulas, quasicopulas and fuzzy logic," *Soft Computing*, vol. 12, no. 12, pp. 1239–1243, 2008.
- [17] C. Sun, Z. Bie, M. Xie, and J. Jiang, "Fuzzy copula model for wind speed correlation and its application in wind curtailment evaluation," *Journal of Renewable Energy*, vol. 93, pp. 68–76, 2016.
- [18] R. R. Yager, "Modeling holistic fuzzy implication using copulas," *Fuzzy Optimization and Decision Making: A Journal of Modeling and Computation Under Uncertainty*, vol. 5, no. 3, pp. 207–226, 2006.




Hindawi

Submit your manuscripts at
www.hindawi.com

