LETTER

Exact Range of the Parameter of an n-Variate FGM Copula under Homogeneous Dependence Structure

Shuhei OTA[†a)], ***Member and*** **Mitsuhiro KIMURA**[††], ***Senior Member***

SUMMARY An n-variate Farlie-Gumbel-Morgenstern (FGM) copula consists of $2^n - n - 1$ parameters that express multivariate dependence among random variables. Motivated by the dependence structure of the n-variate FGM copula, we derive the exact range of the n-variate FGM copula's parameter. The exact range of the parameter is given by a closed-form expression under the condition that all parameters take the same value. Moreover, under the same condition, we reveal that the n-variate FGM copula becomes the independence copula for $n \to \infty$. This result contributes to the dependence modeling such as reliability analysis considering dependent failure occurrence.

key words: *Farlie-Gumbel-Morgenstern copula, exact range, dependence structure, asymptotic property*

1. Introduction

In the reliability analysis of systems, a possible problem arises in the dependent failure of system components. We are liable to underestimate/overestimate the reliability of the system when its components fail dependently. Such a failure is called dependent failure. The dependent failure of the components may occur due to sharing heat, vibration, and tasks [1]. In such a case, we should consider the dependence to precisely assess the reliability of the system. However, most traditional reliability analysis methods assume the independence of the components for its simplicity.

Reliability analysis considering the dependent failure occurrence by using copulas [2], [3] is one of the challenging research topics. The copula is a function that joins several one-dimensional distribution functions to form a multivariate distribution function with dependence. In particular, the Farlie-Gumbel-Morgenstern (FGM, for short) copula is useful as an alternative to a multivariate normal distribution because it has a simple form and can express high-dimensional dependencies among variables. Eryilmaz and Tank [4] investigate the effect of the dependent failure occurrence on the reliability of a system by using the FGM copula.

However, the dependence structure of the FGM copula is a black box. It means that the restrictions of the parameters of the n-variate FGM copula have not been found as closed forms. The n-variate FGM copula proposed by [5] has totally $2^n - n - 1$ dependence parameters which describe the dependencies among the variables (N.B., $\binom{n}{2}$ parameters of them are for the dependencies between any two variables, $\binom{n}{3}$ parameters of them are for among any three variables, and so on). The parameters should be determined so that its joint density function is always non-negative [5]. Although it is easy to confirm whether certain values are acceptable as the parameters of the n-variate FGM copula, the exact range of the parameters of the n-variate FGM copula is unknown. This problem makes the parameter setting in reliability analysis using the n-variate FGM copula difficult.

Motivated by the dependence structure of the n-variate FGM copula, we derive the exact range of the parameter. In this paper, we assume that all parameters are represented by just one parameter θ. In other words, we consider the homogeneous dependence structure of the n-variate FGM copula. Then, we show that the exact range of θ can be represented as a closed-form inequality which depends on n. In addition, we reveal that the n-variate FGM copula becomes the independence copula [2] for $n \to \infty$.

This study contributes to not only the theory of statistics but also dependence modeling using the n-variate FGM copula. The parameter setting and estimation of the n-variate FGM copula become easy if the restriction of the parameters is explicitly given. Besides, we show in which the situation the n-variate FGM copula under the homogeneous dependence structure is suitable as the dependence modeling.

2. Definition

Let $\boldsymbol{U} = (U_1, \ldots, U_n)$ be a random vector that follows an n-variate FGM copula with n uniform marginal distributions in the interval $[0, 1]$. Let $\boldsymbol{\theta}$ denote a set of parameters of the n-variate FGM copula. According to [5], the joint distribution function of the n-variate FGM copula, denoted by C, is defined as

$$\begin{aligned} C(u_1, \ldots, u_n; \boldsymbol{\theta}) &= \Pr[U_1 \le u_1, \ldots, U_n \le u_n] \\ &= \prod_{i=1}^{n} u_i \Bigg(1 + \sum_{1 \le j_1 < j_2 \le n} \theta_{j_1 j_2} (1 - u_{j_1})(1 - u_{j_2}) \\ &\quad + \sum_{1 \le j_1 < j_2 < j_3 \le n} \theta_{j_1 j_2 j_3} (1 - u_{j_1})(1 - u_{j_2})(1 - u_{j_3}) \\ &\quad + \cdots + \theta_{1 \cdots n} (1 - u_1) \cdots (1 - u_n)\Bigg), \end{aligned} \tag{1}$$

where $(u_1, \ldots, u_n) \in [0, 1]^n$ and $\theta_{j_1 \cdots j_k} \in \boldsymbol{\theta}$ is a parameter.

Manuscript received July 27, 2020.
Manuscript revised September 18, 2020.
Manuscript publicized October 27, 2020.

[†]The author is with Faculty of Engineering, Kanagawa Univesity, Yokohama-shi, 221-8686 Japan.
[††]The author is with Faculty of Science & Engineering, Hosei University, Koganei-shi, 184-8584 Japan.
a) E-mail: ota@kanagawa-u.ac.jp

DOI: 10.1587/transfun.2020EAL2080

Equation (1) consists of $2^n - n - 1$ parameters such as $\theta_{12}, \theta_{13}, \ldots, \theta_{123}, \ldots, \theta_{1\cdots n}$. For example, if $n = 3$, we obtain

$$C(u_1, u_2, u_3; \boldsymbol{\theta}) = u_1 u_2 u_3 (1 + \theta_{12}(1-u_1)(1-u_2) + \theta_{13}(1-u_1)(1-u_3) + \theta_{23}(1-u_2)(1-u_3) + \theta_{123}(1-u_1)(1-u_2)(1-u_3)).$$

Moreover, we can write Eq. (1) in a more compact form as follows.

$$C(u_1, \ldots, u_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} u_i \left(1 + \sum_{k=2}^{n} \sum_{1 \le j_1 < \cdots < j_k \le n} \theta_{j_1 \cdots j_k} (1-u_{j_1}) \cdots (1-u_{j_k})\right). \tag{2}$$

The joint density function of $\boldsymbol{U}$ is given by

$$c(u_1, \ldots, u_n; \boldsymbol{\theta}) = \frac{\partial^n}{\partial u_1 \cdots \partial u_n} C(u_1, \ldots, u_n; \boldsymbol{\theta}) = 1 + \sum_{k=2}^{n} \sum_{1 \le j_1 < \cdots < j_k \le n} \theta_{j_1 \cdots j_k} (1-2u_{j_1}) \cdots (1-2u_{j_k}). \tag{3}$$

$\boldsymbol{\theta}$ is the parameter set such that the joint density function is non-negative for every $(u_1, \ldots, u_n) \in [0,1]^n$. Since $c(u_1, \ldots, u_n; \boldsymbol{\theta})$ is a linear function of each u_i $(i = 1, \ldots, n)$, the substitution of $u_i = 0, 1$ in every one of the 2^n possible combinations will yield necessary and sufficient conditions on the values of the elements of $\boldsymbol{\theta}$ (cf. [5]). Thus, $\boldsymbol{\theta}$ must satisfy the following limitation.

$$1 + \sum_{k=2}^{n} \sum_{1 \le j_1 < \cdots < j_k \le n} \theta_{j_1 \cdots j_k} \varepsilon_{j_1} \cdots \varepsilon_{j_k} \ge 0, \tag{4}$$

where $\varepsilon_i = \pm 1$ for $i = 1, \ldots, n$. For example, if $n = 2$, the parameter set becomes $\boldsymbol{\theta} = \{\theta_{12}\}$, and we have $-1 \le \theta_{12} \le 1$. In the case of $n = 3$, the parameters $\boldsymbol{\theta} = \{\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}\}$ are required to hold the following conditions (see [5]).

$$\left.\begin{array}{l} 1 + \theta_{12} + \theta_{13} + \theta_{23} \ge |\theta_{123}| \\ 1 + \theta_{12} - \theta_{13} - \theta_{23} \ge |\theta_{123}| \\ 1 - \theta_{12} + \theta_{13} - \theta_{23} \ge |\theta_{123}| \\ 1 - \theta_{12} - \theta_{13} + \theta_{23} \ge |\theta_{123}| \end{array}\right\}. \tag{5}$$

As a result, Eq. (4) becomes complex as n increases. Therefore, it is hard to understand the dependence structure of the n-variate FGM copula for a large number n unless some specific conditions are made. Although a lot of papers have referred to the restriction for $\boldsymbol{\theta}$ (see e.g. [2], [3], [6] and [7]), few studies explicitly derive the exact ranges of the parameters for $n \ge 4$ in particular because of its complexity. For more details about the FGM copula, see [5] and [8].

3. Literature Review

Several related studies investigated the exact range of $\boldsymbol{\theta}$ of the n-variate FGM copula under certain conditions. Assuming that some parameters are equivalent to 0 and the others are given by a common parameter θ, they derive the range of θ. In order to help the reader's understanding of the exact range of θ, we introduce three well-known results as the propositions shown below.

Johnson and Kotz [5] provided the exact range of $\theta \equiv \theta_{1\cdots n}$ under the condition that all parameters are equal to 0 except for $\theta_{1\cdots n}$ in Eq. (2).

Proposition 3.1: Johnson and Kotz [5]. Suppose the following one-parameter FGM copula.

$$C(u_1, \ldots, u_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} u_i \left(1 + \theta(1-u_1) \cdots (1-u_n)\right). \tag{6}$$

Then, the exact range of θ in Eq. (6) is given by

$$-1 \le \theta \le 1. \tag{7}$$

Mari and Kotz [8] revealed the exact range of θ under the condition that $\theta \equiv \theta_{j_1 j_2}$ for $j_1, j_2 = 1, \ldots, n$ and $j_1 \ne j_2$, and the other $2^n - n - 1 - \binom{n}{2}$ parameters equal 0 in Eq. (2).

Proposition 3.2: Mari and Kotz [8]. Suppose the following one-parameter FGM copula.

$$C(u_1, \ldots, u_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} u_i \left(1 + \theta \sum_{1 \le j_1 < j_2 \le n} (1-u_{j_1})(1-u_{j_2})\right). \tag{8}$$

Then, the exact range of θ in Eq. (8) is given by

$$-\frac{1}{\binom{n}{2}} \le \theta \le \frac{1}{[\frac{n}{2}]}, \tag{9}$$

where $[x]$ denotes the integer part of number x.

Propositions 3.1 and **3.2** provide the exact range of each θ by substituting 0 into most of the original parameters, respectively. However, none of the parameters should be omitted for more generality.

Ota and Kimura [9] found a necessary condition of θ when $\theta \equiv \theta_{j_1 \cdots j_k}$ without omitting any parameters in Eq. (2).

Proposition 3.3: Ota and Kimura [9]. Suppose the following one-parameter FGM copula.

$$C(u_1, \ldots, u_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} u_i \left(1 + \theta \sum_{k=2}^{n} \sum_{1 \le j_1 < \cdots < j_k \le n} (1-u_{j_1}) \cdots (1-u_{j_k})\right). \tag{10}$$

Then, the necessary condition for the exact range of θ in Eq. (10) is given by

$$-\frac{1}{2^n - n - 1} \le \theta \le \frac{1}{(1-v_n)(1+n-2(2-v_n)^{n-1})}. \tag{11}$$

where v_n is uniquely and implicitly defined by the following equation for $0 \leq v_n \leq 1$.

$$n(2 - v_n)^{n-1} - (n-1)(2 - v_n)^{n-2} - \frac{1+n}{2} = 0. \quad (12)$$

Proposition 3.3 supports that θ in Eq. (10) satisfying the condition of Eq. (4) is always in the range of Eq. (11), but not vice versa.

4. Main Results

In this section, we derive the exact range of θ in Eq. (10). We again assume that all parameters of the n-variate FGM copula are equal to a common parameter θ, i.e., $\boldsymbol{\theta} = \{\theta, \ldots, \theta\}$. For example, in the case of $n = 3$, we obtain

$$1 + 3\theta \geq |\theta|, \quad (13)$$

$$1 - \theta \geq |\theta|, \quad (14)$$

from Eq. (5). Then, $-1/4 \leq \theta \leq 1/2$ holds. In general, the exact range of θ is determined by the following theorem.

Theorem 4.1: Suppose the one-parameter FGM copula given by Eq. (10) in which all of the parameters are equivalent to θ. Then, the exact range of θ is given by

$$-\frac{1}{2^n - n - 1} \leq \theta \leq \frac{1}{n-1}. \quad (15)$$

The novelty of **Theorem 4.1** is that the necessary and sufficient condition of θ in Eq. (10) is determined by an explicit formula. This fact makes the parameter setting and estimation of the n-variate FGM copula easy under the homogeneous dependence structure. Note that **Proposition 3.3** shows just a necessary condition of θ in Eq. (10) by an implicit formula.

Figure 1 illustrates the behaviors of the lower and upper bounds of θ given by Eqs. (11) and (15). With respect to Eq. (15), the lower bound is exponentially convergent to 0, and the upper one polynomially goes to 0. Thus, the range of $\theta > 0$ is wider than that of $\theta < 0$. The lower bound of Eq. (15) corresponds to that of Eq. (11), and the upper bound of Eq. (15) is less than that of Eq. (11).

From **Theorem 4.1**, we obtain the following corollary immediately.

Corollary 4.1: For $n \to \infty$, θ in Eq. (10) can only take 0. That is, the n-variate FGM copula under the homogeneous dependence structure asymptotically loses its dependence structure completely.

For $n \to \infty$, the n-variate FGM copula under the homogeneous dependence structure corresponds to the independence copula whose joint distribution function is represented by the product of marginal distributions. Thus, the n-variate FGM copula under the homogeneous dependence structure should be used for the modeling of weak dependence among variables.

We provide the proof of **Theorem 4.1** from the viewpoint of combinatorics. Recall that the exact range of θ is given by Eq. (4). To find the exact range of θ, it is enough to calculate

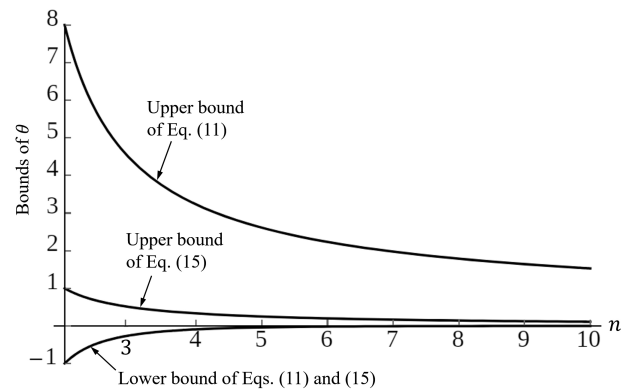

Fig. 1 Behaviors of the lower and upper bounds of θ.

$$\sum_{k=2}^{n} \sum_{1 \leq j_1 < \cdots < j_k \leq n} \varepsilon_{j_1} \cdots \varepsilon_{j_k}, \quad (16)$$

for the following $n + 1$ cases: $(\varepsilon_1, \ldots, \varepsilon_n) = (1, \ldots, 1), (-1, 1, \ldots, 1), \ldots, (-1, \ldots, -1)$ since ε_i and ε_j are symmetric for $i \neq j$ in Eq. (16). Let us define $D(m)$ as

$$D(m) = \sum_{k=2}^{n} \sum_{\substack{1 \leq j_1 < \cdots < j_k \leq n, \\ \varepsilon_1 = \cdots = \varepsilon_m = 1, \\ \varepsilon_{m+1} = \cdots = \varepsilon_n = -1}} \varepsilon_{j_1} \cdots \varepsilon_{j_k}, \quad (17)$$

for $m = 0, 1, \ldots, n$, $\varepsilon_1 = \cdots = \varepsilon_m = 1$ $(0 < m \leq n)$ and $\varepsilon_{m+1} = \cdots = \varepsilon_n = -1$ $(0 \leq m < n)$. Hence, we need to find the exact range of

$$1 + \theta D(m) \geq 0 \quad (18)$$

for $m = 0, 1, \ldots, n$.

The following lemma is used to establish the proof of **Theorem 4.1**.

Lemma 4.1: $D(m)$ can be represented as follows.

$$D(m) = \begin{cases} 2^n - n - 1 & \text{for } m = n \\ n - 1 - 2m & \text{for } 0 \leq m \leq n - 1. \end{cases} \quad (19)$$

Proof of Lemma 4.1: When $m = 0$, i.e., $\varepsilon_1 = \cdots = \varepsilon_n = -1$, we have

$$\begin{aligned} D(0) &= \sum_{1 \leq j_1 < j_2 \leq n} \varepsilon_{j_1}\varepsilon_{j_2} + \sum_{1 \leq j_1 < j_2 < j_3 \leq n} \varepsilon_{j_1}\varepsilon_{j_2}\varepsilon_{j_3} \\ &\qquad + \cdots + \varepsilon_1 \cdots \varepsilon_n \quad (20) \\ &= \sum_{i=2}^{n} (-1)^i \binom{n}{i} \\ &= (1-1)^n + \binom{n}{1} - \binom{n}{0} \\ &= n - 1, \end{aligned}$$

where we use the binomial theorem in the third step.

When $m = 1$, i.e., $\varepsilon_1 = 1$, and $\varepsilon_2 = \cdots = \varepsilon_n = -1$, we separate Eq. (17) into the terms that include ε_1 and those that do not. Then,

$$\begin{aligned} D(1) &= \left(\sum_{1<j_2\le n} \varepsilon_1\varepsilon_{j_2} + \sum_{1<j_2<j_3\le n} \varepsilon_1\varepsilon_{j_2}\varepsilon_{j_3} + \cdots \right) \\ &\quad + \left(\sum_{1<j_1<j_2\le n} \varepsilon_{j_1}\varepsilon_{j_2} + \sum_{1<j_1<j_2<j_3\le n} \varepsilon_{j_1}\varepsilon_{j_2}\varepsilon_{j_3} + \cdots \right) \\ &= \sum_{i=2}^{n} (-1)^{i+1}\binom{n-1}{i-1} + \sum_{i=3}^{n} (-1)^{i+1}\binom{n-1}{i-1} \\ &= n-3. \end{aligned}$$

In the same way, we obtain

$$\begin{aligned} D(2) &= \sum_{i=2}^{n} (-1)^{i+2}\binom{n-2}{i-2} \\ &\quad + \binom{2}{1}\sum_{i=3}^{n} (-1)^{i+2}\binom{n-2}{i-2} + \sum_{i=4}^{n} (-1)^{i+2}\binom{n-2}{i-2}, \\ &\ \ \vdots \end{aligned}$$

$$D(m) = \sum_{h=0}^{m}\binom{m}{h}\sum_{i=2+h}^{n}(-1)^{i+m}\binom{n-m}{i-m}, \tag{21}$$

where $m = 0, 1, \ldots, n$ and $\binom{a}{b} = 0$ for $a < b$ or $b < 0$. After removing all of the terms containing $\binom{n-m}{i-m} = 0$ for $0 \le i \le m$ in Eq. (21), we simplify Eq. (21) as

$$D(m) = \left\{ \sum_{h=0}^{m}\binom{m}{h} \right\} \left\{ \sum_{i=m}^{n}(-1)^{i+m}\binom{n-m}{i-m} \right\} + n - 1 - 2m. \tag{22}$$

Here, we have

$$\sum_{h=0}^{m}\binom{m}{h} = (1+1)^m = 2^m, \tag{23}$$

and

$$\begin{aligned} \sum_{i=m}^{n}(-1)^{i+m}\binom{n-m}{i-m} &= \sum_{i=0}^{n-m}(-1)^{i}\binom{n-m}{i} \\ &= (1-1)^{n-m} \\ &= \begin{cases} 1 & \text{for } m = n \\ 0 & \text{for } 0 \le m \le n-1. \end{cases} \end{aligned} \tag{24}$$

By applying Eqs. (23) and (24) into Eq. (22), we obtain Eq. (19). Hence, the proof is completed. □

Proof of Theory 4.1: Recall Eq. (19). Since

$$\begin{aligned} \min_{0\le m\le n} D(m) &= D(n-1) \\ &= 1-n, \end{aligned}$$

and

$$\begin{aligned} \max_{0\le m\le n} D(m) &= D(n) \\ &= 2^n - n - 1, \end{aligned}$$

hold, Eq. (18) is simplified by

$$-\frac{1}{D(n)} \le \theta \le -\frac{1}{D(n-1)}, \tag{25}$$

where $D(n) > 0$ and $D(n-1) < 0$. Hence, the proof is completed. □

5. Conclusion

In this study, we have revealed the exact range of θ of the n-variate FGM copula in explicit form. The exact range of θ becomes narrow as n increases and is finally convergent to 0 for $n \to \infty$. These facts help to understand the homogeneous dependence structure of the n-variate FGM copula. Moreover, one can say that the n-variate FGM copula under the homogeneous dependence structure corresponds to the independence copula for $n \to \infty$.

On the other hand, the n-variate FGM copula has been used as a modeling tool for high dimensional dependence among data since it has a simple form and can express dependencies among two or more variables. Thus, the results of this study also reveal the limitation of their models to design the dependencies.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP19K23540 and JP19K04892.

References

[1] J. McCool, "Testing for dependency of failure times in life testing," Technometrics, vol.48, no.1, pp.41–48, Jan. 2012.
[2] H. Joe, Dependence Modeling with Copulas, Chapman & Hall, London, 2014.
[3] R.B. Nelsen, An Introduction to Copulas, 2nd ed., Springer-Verlag, Portland, 2006.
[4] S. Eryilmaz and F. Tank, "On reliability analysis of a two-dependent-unit series system with a standby unit," Appl. Math. Comput., vol.218, no.15, pp.7792–7797, April 2012.
[5] N.L. Johnson and S. Kotz, "On some generalized Farlie-Gumbel-Morgenstern distributions," Commun. Stat., vol.4, no.5, pp.415–427, Oct. 1975.
[6] S. Cambanis, "Some properties and generalizations of multivariate Eyraud-Gumbel-Morgenstern distributions," J. Multivariate Anal., vol.7, no.4, pp.551–559, Dec. 1977.
[7] E. Hashorva, "Asymptotic results for FGM random sequences," Statist. Probab. Lett., vol.54, no.4, pp.417–425, Oct. 2001.
[8] D.D. Mari and S. Kotz, Correlation and Dependence, Imperial College Press, London, 2001.
[9] S. Ota and M. Kimura, "On necessary conditions for parameters of minimum and maximum value distributions based on n-variate FGM copula," IEICE Trans. Fundamentals, vol.E102-A, no.3, pp.586–589, March 2019.