LETTER

On Necessary Conditions for Dependence Parameters of Minimum and Maximum Value Distributions Based on n-Variate FGM Copula

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

SUMMARY This paper deals with the minimum and maximum value distributions based on the n-variate FGM copula with one dependence parameter. The ranges of dependence parameters are theoretically determined so that the probability density function always takes a non-negative value. However, the closed-form conditions of the ranges for the dependence parameters have not been known in the literature. In this paper, we newly provide the necessary conditions of the ranges of the dependence parameters for the minimum and maximum value distributions which are derived from FGM copula, and show the asymptotic properties of the ranges.

key words: *minimum value distribution, maximum value distribution, FGM copula, dependence parameter, asymptotic properties*

1. Introduction

Dependence modeling with copulas is one of the challenging research topics of statistical modeling in the last decade. An n-variate copula is a multivariate distribution function having n uniform marginal distributions on the interval [0, 1] (cf., [1] and [2]). Copulas provide powerful tools in the multivariate statistical analysis because one can construct various dependence models by replacing the uniform marginal distributions with other marginal distributions. 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 it can express mutual dependencies among two or more variables. Therefore, the FGM copula has been applied to the statistical modeling in different research fields such as reliability engineering, finance, economics, and so on (e.g., [3] and [4]). For more details about the bivariate FGM copula, see [5] and [6]. In addition, [7] and [8] well explained the n-variate FGM copula.

One unresolved problem of the FGM copula is that restrictions of its parameters have not been found as closed forms. For this reason, for example, it has been difficult to estimate the dependence parameters of the FGM copula so far. The n-variate FGM copula proposed by [7] 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). Note that the parameters should be determined so that its joint density function is always non-negative [7]. In general, a lot of papers have referred to this restriction (e.g., [2], [9], and [10]). However, there are few studies that explicitly derive the exact ranges of the parameters for $n \geq 4$ in particular because of its complexity.

Under such a situation, as a first step, we focus on the minimum and maximum value distributions which are constructed by the n-variate FGM copula in this paper. The former distribution can describe the lifetime (first failure occurrence time) of a series system with n dependent components, and the latter one does that of a dependent parallel system. In addition, in order to reduce the complexity involved in the n-variate FGM copula, we assume that all dependence parameters are represented by just one parameter. As a result, the necessary conditions for the dependence parameter of minimum and maximum distributions are explicitly provided.

In the following section, we explain the fundamental feature of the n-variate FGM copula, and Sect. 3 presents the necessary conditions of the dependence parameters for any given n. The derivation methods of them are presented in Sect. 4. Some discussion on the sufficient conditions and remaining issues on this topic is briefly referred in Sect. 5.

2. Definition

Suppose $\boldsymbol{U} = (U_1, U_2, \ldots, U_n)$ be a random vector that follows an FGM copula with n uniform marginal distributions on the interval [0, 1]. Let C be the joint distribution function of the n-variate FGM copula and let $\mathscr{F}$ be the subsets which consist of all combinations of at least two elements of an index set $\{1, 2, \ldots, n\}$. For example, if $n = 3$, $\mathscr{F} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Denote that $\boldsymbol{\theta}$ is a vector representation of the dependence parameter set of the FGM copula. Then, according to [7], the joint distribution function of $\boldsymbol{U}$ can be written by

$$C(u_1, u_2, \ldots, u_n; \boldsymbol{\theta}) = \Pr[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_n \leq u_n] = \prod_{i=1}^{n} u_i \left(1 + \sum_{S \in \mathscr{F}} \alpha_S \prod_{j \in S} (1 - u_j)\right), \quad (1)$$

where α_S's are the elements of the dependence parameters (i.e., ($\alpha_S \in \boldsymbol{\theta}$)). Moreover, its joint density function is expressed by the following form:

$$c(u_1, u_2, \ldots, u_n; \boldsymbol{\theta})$$

Manuscript received November 7, 2018.

[†]The author is with Graduate School of Science & Engineering, Hosei University, Koganei-shi, 184-8584 Japan.

[††]The author is with Faculty of Science & Engineering, Hosei University, Koganei-shi, 184-8584 Japan.

a) E-mail: kim@hosei.ac.jp

DOI: 10.1587/transfun.E102.A.586

$$= \frac{\partial^n}{\partial u_1 \partial u_2 \cdots \partial u_n} C(u_1, u_2, \ldots, u_n; \boldsymbol{\theta})$$
$$= 1 + \sum_{S \in \mathscr{F}} \alpha_S \prod_{j \in S} (1 - 2u_j). \quad (2)$$

Note that $\boldsymbol{\theta}$ is the parameter set such that the joint density function is non-negative for every $u_i \in [0, 1]$. Thus it is easy to see that α_S's must satisfy the following limitation (cf., [1]).

$$1 + \sum_{S \in \mathscr{F}} \alpha_S \prod_{j \in S} (1 - 2u_j) \geq 0, \quad (3)$$

for $\forall (u_1, u_2, \ldots, u_n) \in [0, 1]^n$. Consequently, for the simplest case, $n = 2$, the parameter set becomes $\boldsymbol{\theta} = \{\alpha_{\{1,2\}}\}$ and we have the well-known range $-1 \leq \alpha_{\{1,2\}} \leq 1$ from Eq. (3).

In the case of $n = 3$, the parameters $\boldsymbol{\theta} = \{\alpha_{\{1,2\}}, \alpha_{\{1,3\}}, \alpha_{\{2,3\}}, \alpha_{\{1,2,3\}}\}$ are required to hold the following conditions (cf., [7] and [12]).

$$\left.\begin{array}{l} 1 + \alpha_{\{1,2\}} + \alpha_{\{1,3\}} + \alpha_{\{2,3\}} \geq |\alpha_{\{1,2,3\}}| \\ 1 + \alpha_{\{1,2\}} - \alpha_{\{1,3\}} - \alpha_{\{2,3\}} \geq |\alpha_{\{1,2,3\}}| \\ 1 - \alpha_{\{1,2\}} + \alpha_{\{1,3\}} - \alpha_{\{2,3\}} \geq |\alpha_{\{1,2,3\}}| \\ 1 - \alpha_{\{1,2\}} - \alpha_{\{1,3\}} + \alpha_{\{2,3\}} \geq |\alpha_{\{1,2,3\}}| \end{array}\right\}. \quad (4)$$

The ranges of these dependence parameters must be determined so as to satisfy Eq. (4). However, it is not so easy to find them. Johnson and Kotz [7] added an additional condition such that $\forall \alpha_S = 0$ for $|S| < n$ in order to loosen the restriction (N.B., $|S|$ means the size of the set S). Then the exact range of $\alpha_{\{1,2,\ldots,n\}}$ can be obtained as

$$-1 \leq \alpha_{\{1,2,\ldots,n\}} \leq 1. \quad (5)$$

As a result, we can see that it is very hard to derive the exact ranges of the dependence parameters unless some certain conditions are made. Therefore in this paper, as we mentioned in Sect. 1, we try to obtain the ranges of the dependence parameters for the minimum and maximum value distributions based on the n-variate FGM copula when all of the dependence parameters are identical to θ. That is, we assume that $\theta \equiv \alpha_S$ for all S in Eq. (1).

3. Main Results

We first show the minimum and maximum distribution functions derived from one-parameter n-variate FGM copula. Let $U_{1:n}$ and $U_{n:n}$ be the minimum and maximum values of $\boldsymbol{U}$, respectively. $C_{1:n}(u; \theta_{1:n})$ and $C_{n:n}(u; \theta_{n:n})$ denote the cumulative distribution functions (CDF) of $U_{1:n}$ and $U_{n:n}$, respectively. According to [11] and [12], $C_{1:n}(u; \theta_{1:n})$ is obtained as

$$C_{1:n}(u; \theta_{1:n}) = \Pr[\min(U_1, U_2, \ldots, U_n) \leq u]$$
$$= 1 - (1 - u)^n \left(1 + \theta_{1:n} \sum_{k=2}^{n} \binom{n}{k} (-u)^k\right). \quad (6)$$

Also $C_{n:n}(u; \theta_{n:n})$ is given by

$$C_{n:n}(u; \theta_{n:n}) = \Pr[\max(U_1, U_2, \ldots, U_n) \leq u]$$
$$= u^n \left(1 + \theta_{n:n} \sum_{k=2}^{n} \binom{n}{k} (1 - u)^k\right). \quad (7)$$

Note that we rewrite θ to $\theta_{1:n}$ and $\theta_{n:n}$ in the above equations respectively in order to distinguish these two parameters.

The following theorems present the necessary conditions for the ranges of $\theta_{1:n}$ and $\theta_{n:n}$ in Eqs. (6) and (7), respectively. In addition, these theorems yield corollaries about their asymptotic properties.

Theorem 3.1: The range of $\theta_{1:n}$ is given by the following inequality.

$$-\frac{1}{n-1} \leq \theta_{1:n} \leq \frac{1}{2 - 2(1 - u_n^*)^n - (1 + n)u_n^*}, \quad (8)$$

where

$$u_n^* = 1 - \left(\frac{1 + n}{2n}\right)^{\frac{1}{n-1}}. \quad (9)$$

Corollary 3.1: As $n \to \infty$, the range of $\theta_{1:n}$ is obtained as

$$0 \leq \theta_{1:n} \leq \frac{1}{1 - \log 2} \simeq 3.259. \quad (10)$$

Theorem 3.2: The range of $\theta_{n:n}$ is given by the following inequality.

$$-\frac{1}{2^n - 1 - n} \leq \theta_{n:n} \leq \frac{1}{(1 - v_n^*)\{1 + n - 2(2 - v_n^*)^{n-1}\}}, \quad (11)$$

where v_n^* is uniquely defined by the solution of the following equation.

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

for $0 \leq v_n^* \leq 1$.

Remark 3.1: The asymptotic property of v_n^* is as follows.

$$\lim_{n \to \infty} v_n^* = 1. \quad (13)$$

This remark offers some support to the next conjecture.

Conjecture 3.1: As $n \to \infty$, the following equation holds.

$$\lim_{n \to \infty} \frac{1}{(1 - v_n^*)\{1 + n - 2(2 - v_n^*)^{n-1}\}} = 0. \quad (14)$$

This conjecture is derived by **Remark 3.1** and the assumption that $1 + n - 2(2 - v_n^*)^{n-1}$ diverges to positive infinity more quickly than $1 - v_n^*$ converges to 0. Then we have the following corollary.

Corollary 3.2: As $n \to \infty$, the range of $\theta_{n:n}$ is convergent to 0.

Table 1 Numerical results of the ranges of $\theta_{1:n}$ and $\theta_{n:n}$.

n	$\theta_{1:n}$	$\theta_{n:n}$	n	$\theta_{1:n}$	$\theta_{n:n}$
2	[−1.000, 8.000]	[−1.000, 8.000]	17	[−0.063, 3.559]	[−0.000, 1.104]
3	[−0.500, 5.639]	[−0.250, 4.440]	18	[−0.059, 3.541]	[−0.000, 1.070]
4	[−0.333, 4.850]	[−0.091, 3.229]	19	[−0.056, 3.526]	[−0.000, 1.039]
5	[−0.250, 4.454]	[−0.038, 2.611]	20	[−0.053, 3.512]	[−0.000, 1.011]
6	[−0.200, 4.216]	[−0.018, 2.232]	21	[−0.050, 3.499]	[−0.000, 0.985]
7	[−0.167, 4.057]	[−0.008, 1.974]	22	[−0.048, 3.488]	[−0.000, 0.961]
8	[−0.143, 3.943]	[−0.004, 1.786]	23	[−0.045, 3.477]	[−0.000, 0.940]
9	[−0.125, 3.858]	[−0.002, 1.642]	24	[−0.043, 3.468]	[−0.000, 0.920]
10	[−0.111, 3.792]	[−0.001, 1.528]	25	[−0.042, 3.459]	[−0.000, 0.901]
11	[−0.100, 3.738]	[−0.000, 1.435]	26	[−0.040, 3.451]	[−0.000, 0.884]
12	[−0.091, 3.695]	[−0.000, 1.357]	27	[−0.038, 3.444]	[−0.000, 0.868]
13	[−0.083, 3.659]	[−0.000, 1.292]	28	[−0.037, 3.437]	[−0.000, 0.852]
14	[−0.077, 3.628]	[−0.000, 1.235]	29	[−0.036, 3.430]	[−0.000, 0.838]
15	[−0.071, 3.602]	[−0.000, 1.186]	30	[−0.034, 3.425]	[−0.000, 0.825]
16	[−0.067, 3.579]	[−0.000, 1.143]	∞	[0, 1/(1 − log2)]	0

In summary, the above results guarantee that the ranges of $\theta_{1:n}$ and $\theta_{n:n}$ depend on only n, and the ranges become narrower as n increases. For example, Table 1 shows the numerical results of the ranges of $\theta_{1:n}$ and $\theta_{n:n}$ for $n = 2, 3, \ldots, 30$, and ∞. Note that every value is calculated by Eqs. (8) and (11), and rounded to the nearest thousandth. This table implies that the ranges of the negative values of $\theta_{1:n}$ and $\theta_{n:n}$ are narrower than those of the positive values, respectively. In addition, we can find that the range of $\theta_{n:n}$ is narrower than that of $\theta_{1:n}$.

4. Proofs

Proof of Theorem 3.1: We need to derive the closed form of the limitation $\theta_{1:n} \in \{\theta \mid \frac{d}{du}C_{1:n}(u;\theta) \geq 0, 0 \leq u \leq 1\}$. First we define the probability density function (PDF) of $U_{1:n}$ by the following $c_{1:n}(u;\theta_{1:n})$.

$$\begin{aligned} c_{1:n}(u;\theta_{1:n}) &= \frac{d}{du}C_{1:n}(u;\theta_{1:n}) \\ &= n(1-u)^{n-1}\left\{1 + \theta_{1:n}(-2 + 2(1-u)^n + (1+n)u)\right\}. \end{aligned} \tag{15}$$

Here, $\theta_{1:n}$ has the following relationship.

$$\begin{aligned} &\theta_{1:n} \in \{\theta \mid c_{1:n}(u;\theta) \geq 0, 0 \leq u \leq 1\} \\ &\Leftrightarrow \theta_{1:n} \in \{\theta \mid \min_{0 \leq u \leq 1}[c_{1:n}(u;\theta)] \geq 0\}. \end{aligned} \tag{16}$$

That is, the problem is equivalent to finding $\theta_{1:n}$ such that the minimum value of $c_{1:n}(u;\theta_{1:n})$ is non-negative. Since $n(1-u)^{n-1} \geq 0$ for $0 \leq u \leq 1$ and $n \geq 2$, we have

$$\theta_{1:n} \in \{\theta \mid \min_{0 \leq u \leq 1}[\tilde{c}_{1:n}(u;\theta)] \geq 0\}, \tag{17}$$

where

$$\tilde{c}_{1:n}(u;\theta_{1:n}) = \frac{1}{n(1-u)^{n-1}}c_{1:n}(u;\theta_{1:n}). \tag{18}$$

Hence, we can verify the **Theorem 3.1** by solving the minimization problem of Eq. (17). In order to solve this problem, we define the first and second derivatives with respect to u of $\tilde{c}_{1:n}(u;\theta_{1:n})$ as follows.

$$\begin{aligned} \tilde{c}'_{1:n}(u;\theta_{1:n}) &\overset{\text{def}}{=} \frac{d}{du}\tilde{c}_{1:n}(u;\theta_{1:n}) \\ &= \theta_{1:n}(-2n(1-u)^{n-1} + 1 + n), \end{aligned} \tag{19}$$

$$\begin{aligned} \tilde{c}''_{1:n}(u;\theta_{1:n}) &\overset{\text{def}}{=} \frac{d^2}{du^2}\tilde{c}_{1:n}(u;\theta_{1:n}) \\ &= \theta_{1:n}(n-1)n(1-u)^{n-2}. \end{aligned} \tag{20}$$

Let u_n^* be an critical point of $\tilde{c}_{1:n}$ (i.e., $\tilde{c}'_{1:n}(u_n^*;\theta_{1:n}) = 0$). Then, u_n^* uniquely exists on the interval [0, 1], and it is easy to see that $u_n^* = 1 - \left(\frac{1+n}{2n}\right)^{\frac{1}{n-1}}$.

Consider $\theta_{1:n} \geq 0$. In this case, for $0 \leq u \leq 1$, $\tilde{c}_{1:n}(u;\theta_{1:n})$ is a convex function because $\tilde{c}''_{1:n}(u;\theta_{1:n}) \geq 0$. Thus, $\tilde{c}_{1:n}(u_n^*;\theta_{1:n})$ is the absolute minimum value. Hence, we have $\theta_{1:n} \in \{\theta \mid \tilde{c}_{1:n}(u_n^*;\theta) \geq 0\}$. This yields

$$\theta_{1:n} \leq \frac{1}{2 - 2(1-u_n^*)^n - (1+n)u_n^*}, \tag{21}$$

where $1/(2 - 2(1-u_n^*)^n - (1+n)u_n^*)$ gives the upper bound of $\theta_{1:n}$.

Consider $\theta_{1:n} < 0$. In this case, $\tilde{c}_{1:n}(u;\theta_{1:n})$ is a concave function that takes the minimum value if and only if $u = 1$. Thus, we have $\theta_{1:n} \in \{\theta \mid \tilde{c}_{1:n}(1;\theta) \geq 0\}$. This implies

$$-\frac{1}{n-1} \leq \theta_{1:n}, \tag{22}$$

where $-1/(n-1)$ gives the lower bound of $\theta_{1:n}$. Hence, the proof is complete. □

Proof of Corollary 3.1: By **Theorem 3.1**, the upper bound of $\theta_{1:n}$ is given by $1/(2 - 2(1-u_n^*)^n - (1+n)u_n^*)$. Thus as $n \to \infty$, the upper bound of $\theta_{1:n}$ can be written by

$$\lim_{n\to\infty} \frac{1}{2 - 2(\frac{1+n}{2n})^{\frac{n}{n-1}} - (1+n)(1 - \frac{1+n}{2n})^{\frac{1}{n-1}}}. \tag{23}$$

Note that u_n^* is replaced by Eq. (9). By considering its Taylor series, Eq. (23) equals to

$$\lim_{n\to\infty} \frac{1}{2-(1+O(\frac{1}{n}))-(\log 2+O(\frac{1}{n}))} = \frac{1}{1-\log 2}, \tag{24}$$

where $O(\cdot)$ denotes Landau's symbol.

Moreover, for the lower bound, we have

$$\lim_{n\to\infty} -\frac{1}{n-1} = 0. \tag{25}$$

Hence, the proof is complete. □

We would like to omit the proofs of **Theorem 3.2** and **Corollary 3.2** because they can be shown in the same way as those of **Theorem 3.1** and **Corollary 3.1**.

5. Discussion and Concluding Remarks

The results in the previous section have been obtained from the necessary condition such that Eq. (15) always takes non-negative value for any $u \in [0,1]$ and n in the case of the minimum value distribution (and the maximum one as well). Therefore the ranges obtained by Eqs. (8) and (11) are both lack of sufficiency. In order to show this fact, let us go back to the FGM copula presented in Eq. (1) with $n = 3$. By setting all the parameters $\alpha_{\{1,2\}}$, $\alpha_{\{1,3\}}$, $\alpha_{\{2,3\}}$, and $\alpha_{\{1,2,3\}}$ be identical to θ in Eq. (4), we can calculate the possible range of θ as

$$-\frac{1}{4} \le \theta \le \frac{1}{2}. \tag{26}$$

On the other hand, from Table 1, we recall

$$-\frac{1}{2} \le \theta_{1:3} \le 5.639, \tag{27}$$

$$-\frac{1}{4} \le \theta_{3:3} \le 4.440, \tag{28}$$

for the two kinds of dependence parameters, respectively. It should be noted that the minimum and maximum value distributions which have been dealt with in this paper both exist only if the n-variate FGM copula is theoretically valid. In other words, for example, if we estimate that the value of $\hat{\theta}_{1:3}$ is 2.0 from a certain data analysis concerning the minimum value distribution, this value surely satisfies Eqs. (8) and (27) but it does not satisfy the limitation denoted by Eq. (3) with the identical dependence parameters. Therefore in this case, we must discard the estimation result $\hat{\theta}_{1:3} = 2.0$, and need to find the estimated value from the range $-\frac{1}{4} \le \theta \le \frac{1}{2}$ in Eq. (26) instead of Eq. (27).

In conclusion, we have revealed the necessary conditions for the dependence parameters $\theta_{1:n}$ and $\theta_{n:n}$ for general n and their asymptotic properties on $n \to \infty$ in this paper. However, providing their sufficient conditions and the exact ranges of the parameters has been still remaining for the future work.

References

[1] P. Jaworski, F. Durante, W. Härdle, and T. Rychlik, eds., Copula Theory and Its Applications, Springer-Verlag, Berlin, 2010.

[2] R.B. Nelsen, An Introduction to Copulas, second ed., Springer-Verlag, Portland, 2010.

[3] 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.

[4] H. Cossette, M.P. Cote, E. Marceau, and K. Moutanabbir, "Multivariate distribution defined with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation," Insurance Math. Econom, vol.52, no.3, pp.560–572, May 2013.

[5] D.J. Farlie, "The performance of some correlation coefficients for a general bivariate distribution," Biometrika, vol.47, no.3-4, pp.307–323, Dec. 1960.

[6] E.J. Gumbel, "Bivariate exponential distributions," J. Am. Stat. Assoc., vol.55, no.292, pp.698–707, Dec. 1960.

[7] N.L. Johnson and S. Kotz, "On some generalized Farlie-Gumbel-Morgenstern distributions," Commun. Stat., vol.4, no.5, pp.415–427, Oct. 1975.

[8] D.D. Mari and S. Kotz, Correlation and Dependence, Imperial College Press, London, 2001.

[9] S. Cambanis, "Some properties and generalizations of multivariate Eyraud-Gumbel-Morgenstern distributions," J. Multivariate Annal., vol.7, no.4, pp.551–559, Dec. 1977.

[10] E. Hashorva, "Asymptotic results for FGM random sequences," Statist. Probab. Lett., vol.54, no.4, pp.417–425, Oct. 2001.

[11] J. Navarro, J.M. Ruiz, and C.J. Sandoval, "Properties of coherent systems with dependent components," Commun. Stat., vol.36, no.1, pp.175–191, Feb. 2007.

[12] M. Kimura, S. Ota, and S. Abe, "Some reliability properties of n-component parallel/series systems with interdependent failures," in Reliability Modeling with Computer and Maintenance Applications, S. Nakamura, C.H. Qian, and T. Nakagawa, eds., pp.155–173, World Scientific, Singapore, 2017.