



# Effective estimation algorithm for parameters of multivariate Farlie–Gumbel–Morgenstern copula

Shuhei Ota<sup>1</sup> · Mitsuhiro Kimura<sup>2</sup>

Received: 13 December 2020 / Accepted: 24 March 2021 / Published online: 19 April 2021  
© The Author(s) 2021

## Abstract

This paper focuses on the parameter estimation for the  $d$ -variate Farlie–Gumbel–Morgenstern (FGM) copula ( $d \geq 2$ ), which has  $2^d - d - 1$  dependence parameters to be estimated; therefore, maximum likelihood estimation is not practical for a large  $d$  from the viewpoint of computational complexity. Besides, the restriction for the FGM copula's parameters becomes increasingly complex as  $d$  becomes large, which makes parameter estimation difficult. We propose an effective estimation algorithm for the  $d$ -variate FGM copula by using the method of inference functions for margins under the restriction of the parameters. We then discuss its asymptotic normality as well as its performance determined through simulation studies. The proposed method is also applied to real data analysis of bearing reliability.

**Keywords** Inference functions for margins · Maximum likelihood estimation · FGM copula · Asymptotic normality · Multivariate distribution

## 1 Introduction

A copula is a function that joins several one-dimensional distribution functions to form a multivariate distribution function with dependency (Nelsen 2006). The copula works as a multivariate distribution function by itself.

---

This work was supported by JSPS KAKENHI Grant Numbers JP19K23540, JP19K04892, JP21K14373.

✉ Shuhei Ota  
ota@kanagawa-u.ac.jp

Mitsuhiro Kimura  
kim@hosei.ac.jp

<sup>1</sup> Department of Industrial Engineering and Management, Kanagawa University, 3-27-1, Rokkakubashi, Kanagawa-ku, Yokohama-shi, Kanagawa 221-8686, Japan

<sup>2</sup> Department of Industrial and Systems Engineering, Hosei University, 3-7-2, Kajino-cho, Koganei-shi, Tokyo 184-8584, Japan

In reliability engineering, the copula is used as a useful tool for reliability analysis of systems in which dependent failures sometimes occur. A dependent failure of systems and/or system components may occur due to sharing heat, vibration, and tasks (McCool 2012). We should consider the dependence of components to precisely assess the reliability of the system. For example, Eryilmaz and Tank (2012) showed that the positive dependence between two components improves the mean time to failure of a two-component series system by using an Farlie–Gumbel–Morgenstern (FGM) copula and Ali–Mikhail–Haq copula. Navarro et al. (2007) developed a reliability model for coherent systems with dependent components by using several types of copulas including the FGM copula.

There have been many studies on the FGM copula. Earlier studies (Farlie 1960; Gumbel 1960; Morgenstern 1956) discussed families of the bivariate FGM copula. Johnson and Kotz (1975) formulated the FGM copula as a multivariate distribution. The multivariate FGM copula is useful as an alternative to a multivariate normal distribution because it has a simple form and can express mutual dependencies among two or more variables (Jaworski et al. 2010). Therefore, the FGM copula has been applied to statistical modeling in various research fields such as economics (Patton 2006), educational engineering (Shih et al. 2019), finance (Cossette et al. 2013) and reliability engineering (Navarro et al. 2007; Navarro and Durante 2017).

However, there are no practical estimation methods of the parameters of the multivariate FGM copula. The reason mainly depends on the computational complexity of estimating a large number of the parameters. It is known that the  $d$ -variate FGM copula consists of  $2^d - d - 1$  constrained parameters for  $d \geq 2$ . In fact, the  $d$ -variate FGM copula has relatively many parameters compared with other multivariate copulas (e.g., a  $d$ -variate Gaussian copula has  $d(d-1)/2 (\leq 2^d - d - 1)$  correlation parameters (Jaworski et al. 2010)). This implies that estimation for the multivariate FGM copula's parameters is computationally more complicated than that for the other copulas' parameters. Besides, maximum likelihood estimation (MLE) is infeasible over such a high dimensional space with the parameter constraint.

The method of inference functions for margins (IFM) is one acceptable estimation method for parameters of copulas. It was proposed by Joe and Xu (1996), and extensively reviewed by Xu (1996). This method first estimates the univariate parameters from separate univariate likelihood functions of a copula. It then estimates bivariate, trivariate and multivariate parameters step by step from respective bivariate, trivariate, and multivariate marginal likelihoods with lower order parameters fixed as the estimated values. The advantages of IFM are its computational efficiency and asymptotic normality that holds under several regularity conditions (Xu 1996; Joe 2005; Patton 2006). However, the following two problems exist when we apply IFM to the multivariate FGM copula: (1) it is not known how to consider the complex restriction for the FGM copula's parameters, and (2) it has not been found whether the estimators of IFM for the multivariate FGM copula have asymptotic normality.

In this paper, we propose an effective estimation algorithm for the multivariate FGM copula by using IFM. The proposed algorithm requires linear optimization to consider the restriction of the multivariate FGM copula's parameters. Besides, we analyze the asymptotic normality of the estimators given by the proposed algorithm.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the definition and property of the multivariate FGM copula. In Sect. 3, we propose estimation algorithms for the multivariate FGM copula based on MLE and IFM. Their asymptotic properties are analytically investigated in Sect. 4. Section 5 consists of simulation results to compare MLE and IFM for the multivariate FGM copula from the viewpoint of estimation accuracy and computation time. In Sect. 6, our proposed method is applied to real data analysis of bearing reliability. Finally, we conclude our study with a summary in Sect. 7.

## 2 FGM copula

Let  $U = (U_1, \dots, U_d)$  be a random vector that follows a  $d$ -variate FGM copula with  $d$  uniform marginal distributions in the interval  $[0, 1]$ . Let  $\theta$  denote the parameter vector of the  $d$ -variate FGM copula. According to Johnson and Kotz (1975) and p. 108 of Nelsen (2006), the joint distribution function of the  $d$ -variate FGM copula, denoted by  $C_d$ , is defined as

$$C_d(u_1, \dots, u_d; \theta) = \Pr[U_1 \leq u_1, \dots, U_d \leq u_d] \\ = \prod_{j=1}^d u_j \left( 1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} (1 - u_{j_1}) \cdots (1 - u_{j_k}) \right),$$

where  $(u_1, \dots, u_d) \in [0, 1]^d$  and  $\theta_{j_1 \dots j_k} \in \theta$  is a parameter. Since  $\theta \neq \mathbf{0}$  means that the random variables are obviously dependent, we call an element of  $\theta$  a dependence parameter. For  $d = 2$  and 3, we obtain

$$C_2(u_1, u_2; \theta) = u_1 u_2 (1 + \theta_{12} (1 - u_1) (1 - u_2)), \\ C_3(u_1, u_2, u_3; \theta) = u_1 u_2 u_3 (1 + \theta_{12} (1 - u_1) (1 - u_2) + \theta_{13} (1 - u_1) (1 - u_3) \\ + \theta_{23} (1 - u_2) (1 - u_3) + \theta_{123} (1 - u_1) (1 - u_2) (1 - u_3)).$$

The  $d$ -variate FGM copula consists of  $\sum_{j=2}^d \binom{d}{j} = 2^d - d - 1$  dependence parameters. The correlation and regression properties for any pair of  $U$  were studied by Johnson and Kotz (1977).

The joint density function of  $U$ , denoted by  $c_d$ , is given by

$$c_d(u_1, \dots, u_d; \theta) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \dots, u_d; \theta) \\ = 1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} (1 - 2u_{j_1}) \cdots (1 - 2u_{j_k}).$$

$\theta$  is a parameter vector such that the joint density function  $c_d(u_1, \dots, u_d; \theta)$  is non-negative for every  $u_j \in [0, 1]$ . Thus,  $\theta$  must satisfy the following restriction (Johnson and Kotz 1975; Jaworski et al. 2010).

$$1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} (1 - 2u_{j_1}) \dots (1 - 2u_{j_k}) \geq 0, \quad u_j \in [0, 1], \tag{1}$$

for  $j = 1, \dots, d$ . Since  $c_d(u_1, \dots, u_d; \theta)$  is a linear function of each  $u_j$ , the substitution of  $u_j = 0, 1$  in every  $2^d$  possible combinations will yield necessary and sufficient conditions of  $\theta$  (Johnson and Kotz 1975). Thus, we can write Eq. (1) in a compact form as follows.

$$1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} \varepsilon_{j_1} \dots \varepsilon_{j_k} \geq 0, \quad \varepsilon_j = \pm 1, \tag{2}$$

for  $j = 1, \dots, d$ . For example, when  $d = 2$ , the parameter vector becomes  $\theta = (\theta_{12})$ , and we have  $-1 \leq \theta_{12} \leq 1$ . When  $d = 3$ , the parameters  $\theta = (\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123})$  are required to hold  $2^3$  inequalities as follows (Johnson and Kotz 1975).

$$\left. \begin{aligned} 1 + \theta_{12} + \theta_{13} + \theta_{23} + \theta_{123} &\geq 0 & (\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 1) \\ 1 + \theta_{12} - \theta_{13} - \theta_{23} - \theta_{123} &\geq 0 & (\varepsilon_1 = \varepsilon_2 = 1, \varepsilon_3 = -1) \\ 1 - \theta_{12} + \theta_{13} - \theta_{23} - \theta_{123} &\geq 0 & (\varepsilon_1 = \varepsilon_3 = 1, \varepsilon_2 = -1) \\ 1 - \theta_{12} - \theta_{13} + \theta_{23} - \theta_{123} &\geq 0 & (\varepsilon_2 = \varepsilon_3 = 1, \varepsilon_1 = -1) \\ 1 - \theta_{12} - \theta_{13} + \theta_{23} + \theta_{123} &\geq 0 & (\varepsilon_1 = 1, \varepsilon_2 = \varepsilon_3 = -1) \\ 1 - \theta_{12} + \theta_{13} - \theta_{23} + \theta_{123} &\geq 0 & (\varepsilon_2 = 1, \varepsilon_1 = \varepsilon_3 = -1) \\ 1 + \theta_{12} - \theta_{13} - \theta_{23} + \theta_{123} &\geq 0 & (\varepsilon_3 = 1, \varepsilon_1 = \varepsilon_2 = -1) \\ 1 + \theta_{12} + \theta_{13} + \theta_{23} - \theta_{123} &\geq 0 & (\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = -1) \end{aligned} \right\} \tag{3}$$

The condition given by Eq. (2) becomes increasingly complex as  $d$  becomes large. When we estimate the dependence parameters, we need to consider this condition.

Multivariate FGM distributions can be constructed by the Sklar’s theorem (Nelsen 2006). Suppose  $X = (X_1, \dots, X_d)$  is a random vector that follows the  $d$ -variate FGM distribution with arbitrary continuous marginal distributions. For  $j = 1, \dots, d$ , let  $F_j(x_j; \delta_j)$  and  $f_j(x_j; \delta_j)$  be the  $j$ th marginal distribution function and the density function with a parameter vector  $\delta_j$ , respectively. Let us define  $\delta$  as  $\delta = (\delta_1, \dots, \delta_d)$ . Then, the cumulative distribution function of  $X$ , denoted by  $H_d$ , is represented as

$$\begin{aligned} H_d(x_1, \dots, x_d; \delta, \theta) &= \Pr[X_1 \leq x_1, \dots, X_d \leq x_d] \\ &= C_d(F_1(x_1; \delta_1), \dots, F_d(x_d; \delta_d); \theta) \\ &= \prod_{j=1}^d F_j(x_j; \delta_j) \left( 1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} \bar{F}_{j_1}(x_{j_1}; \delta_{j_1}) \dots \bar{F}_{j_k}(x_{j_k}; \delta_{j_k}) \right), \end{aligned}$$

where  $\bar{F}_j(x_j; \delta_j) \equiv 1 - F_j(x_j; \delta_j)$ .

The joint density function of  $X$ , denoted by  $h_d$ , is given by

$$\begin{aligned}
 h_d(x_1, \dots, x_d; \boldsymbol{\delta}, \boldsymbol{\theta}) &= \frac{\partial^d}{\partial x_1 \dots \partial x_d} H_d(x_1, \dots, x_d; \boldsymbol{\delta}, \boldsymbol{\theta}) \\
 &= \prod_{j=1}^d f_j \left( 1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} (1 - 2F_{j_1}) \dots (1 - 2F_{j_k}) \right),
 \end{aligned}$$

where  $f_j \equiv f_j(x_j; \boldsymbol{\delta}_j)$  and  $F_j \equiv F_j(x_j; \boldsymbol{\delta}_j)$ .

### 3 Estimation

In this section, we first introduce MLE and its limitation for the multivariate FGM distribution. To solve the limitation, we then propose an estimation algorithm for the multivariate FGM distribution by using IFM.

#### 3.1 Maximum likelihood estimation

MLE is one of the natural choices to estimate parameters of probability distributions because its estimators satisfy asymptotic normality under several regularity conditions (Lehmann and Casella 1998; Joe 2014). With this method, we estimate all the parameters  $\boldsymbol{\delta}$  and  $\boldsymbol{\theta}$  simultaneously. For a sample size  $n$ , with observed independent random vectors  $\mathbf{x}_i = (x_{i1}, \dots, x_{id})$  for  $i = 1, \dots, n$ , the full-dimensional log-likelihood function of  $\boldsymbol{\delta}$  and  $\boldsymbol{\theta}$  can be written as

$$\begin{aligned}
 \ell &\equiv \ell(\boldsymbol{\delta}, \boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) \\
 &\equiv \sum_{i=1}^n \log h_d(x_{i1}, \dots, x_{id}; \boldsymbol{\delta}, \boldsymbol{\theta}) \\
 &= \sum_{i=1}^n \sum_{j=1}^d \log f_j + \sum_{i=1}^n \log \left( 1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1 \dots j_k} (1 - 2F_{j_1}) \dots (1 - 2F_{j_k}) \right).
 \end{aligned}$$

Thus, the estimators of MLE, denoted as  $\hat{\boldsymbol{\delta}}$  and  $\hat{\boldsymbol{\theta}}$ , are given by

$$(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}) = \arg \max_{\boldsymbol{\delta}, \boldsymbol{\theta}} \ell(\boldsymbol{\delta}, \boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n). \tag{4}$$

To find  $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}})$ , we need to use numerical optimization techniques because Eq. (4) does not have closed form expression. The problems of numerical optimization in Eq. (4) are high-dimensionality and the parameter constraints given by Eq. (2). Numerical optimization in a high-dimensional and constrained space requires a large amount of computational resources. This issue relates to computational feasibility.

Numerical optimization might take a large amount of computation time and fail as the dimension increases. Also, the estimation accuracy of MLE worsens because the numerical optimization technique may not find the global optimum but a local one. Even if all the marginal distributions are uni-parameter distributions (i.e.,  $\boldsymbol{\delta}_j = \delta_j$ ), Eq. (4) consists of the  $2^d - 1$  unknown parameters  $\boldsymbol{\delta}$  and  $\boldsymbol{\theta}$ . Moreover, the ignorance of

parameter constraints is one of the most common reasons why the Newton-Raphson type algorithms diverge (MacDonald 2014). Therefore, other simple estimation methods have been required for the multivariate FGM distribution’s parameters.

### 3.2 The method of inference functions for margins

To reduce the computational difficulty of MLE, we introduce IFM for estimation of the multivariate FGM distribution. With this method, we estimate the dependence parameters one by one so that each estimation result satisfies the condition given by Eq. (2).

As the first step, let us define  $k$ -dimensional marginal likelihood functions for  $k = 1, 2, \dots, d - 1$ . It is easy to see that the  $k$ -dimensional marginal distribution function is as follows.

$$H_k(x_1, \dots, x_k; \cdot) = \prod_{j=1}^k F_j \left( 1 + \sum_{p=2}^k \sum_{1 \leq j_1 < \dots < j_p \leq k} \theta_{j_1 \dots j_p} (1 - F_{j_1}) \dots (1 - F_{j_p}) \right).$$

Thus, for example, we have

$$\begin{aligned} H_1(x_{j_1}; \delta_{j_1}) &= F_{j_1}, \\ H_2(x_{j_1}, x_{j_2}; \delta_{j_1}, \delta_{j_2}, \theta_{j_1 j_2}) &= F_{j_1} F_{j_2} (1 + \theta_{j_1 j_2} (1 - F_{j_1})(1 - F_{j_2})), \\ H_3(x_{j_1}, x_{j_2}, x_{j_3}; \delta_{j_1}, \delta_{j_2}, \delta_{j_3}, \theta_{j_1 j_2}, \theta_{j_1 j_3}, \theta_{j_2 j_3}, \theta_{j_1 j_2 j_3}) \\ &= F_{j_1} F_{j_2} F_{j_3} (1 + \theta_{j_1 j_2} (1 - F_{j_1})(1 - F_{j_2}) + \theta_{j_1 j_3} (1 - F_{j_1})(1 - F_{j_3}) \\ &\quad + \theta_{j_2 j_3} (1 - F_{j_2})(1 - F_{j_3}) + \theta_{j_1 j_2 j_3} (1 - F_{j_1})(1 - F_{j_2})(1 - F_{j_3})). \end{aligned}$$

The key point of these functions is that their variables are controlled by only the parameters with the same subscripts of the marginal distribution functions. We then obtain the following log-likelihood functions of the  $k$ -dimensional marginal distribution for  $k < d$  and  $1 \leq j_1 < \dots < j_k \leq d$ .

$$\begin{aligned} \ell_{j_1} &\equiv \ell_{j_1}(\delta_{j_1}; \mathbf{x}_1, \dots, \mathbf{x}_n) \equiv \sum_{i=1}^n \log f_{j_1}(x_{ij_1}; \delta_{j_1}), \\ \ell_{j_1 j_2} &\equiv \ell_{j_1 j_2}(\delta_{j_1}, \delta_{j_2}, \theta_{j_1 j_2}; \mathbf{x}_1, \dots, \mathbf{x}_n) \\ &\equiv \sum_{i=1}^n \log h_2(x_{ij_1}, x_{ij_2}; \delta_{j_1}, \delta_{j_2}, \theta_{j_1 j_2}), \\ \ell_{j_1 j_2 j_3} &\equiv \ell_{j_1 j_2 j_3}(\delta_{j_1}, \delta_{j_2}, \delta_{j_3}, \theta_{j_1 j_2}, \theta_{j_1 j_3}, \theta_{j_2 j_3}, \theta_{j_1 j_2 j_3}; \mathbf{x}_1, \dots, \mathbf{x}_n) \\ &\equiv \sum_{i=1}^n \log h_3(x_{ij_1}, x_{ij_2}, x_{ij_3}; \delta_{j_1}, \delta_{j_2}, \delta_{j_3}, \theta_{j_1 j_2}, \theta_{j_1 j_3}, \theta_{j_2 j_3}, \theta_{j_1 j_2 j_3}), \\ &\vdots \\ \ell_{j_1 \dots j_{d-1}} &\equiv \ell_{j_1 \dots j_{d-1}}(\delta_{j_1}, \dots, \delta_{j_{d-1}}, \theta_{j_1 j_2}, \dots, \theta_{j_1 \dots j_{d-1}}; \mathbf{x}_1, \dots, \mathbf{x}_n) \\ &\equiv \sum_{i=1}^n \log h_{d-1}(x_{ij_1}, \dots, x_{ij_{d-1}}; \delta_{j_1}, \dots, \delta_{j_{d-1}}, \theta_{j_1 j_2}, \dots, \theta_{j_1 \dots j_{d-1}}). \end{aligned}$$

Now, we propose an estimation algorithm based on IFM for the  $d$ -variate FGM distribution. With IFM, we first estimate the parameters of marginal distributions as

$$\tilde{\delta}_{j_1} = \operatorname{argmax}_{\delta_{j_1}} \ell_{j_1}(\delta_{j_1}; \mathbf{x}_{j_1}),$$

for  $j_1 = 1, 2, \dots, d$ .

Next, we estimate bivariate, trivariate and  $d$ -variate dependence parameters step by step. For  $1 \leq j_1 < j_2 \leq d$ , the bivariate dependence parameters  $\theta_{j_1 j_2}$  are estimated as follows.

$$\tilde{\theta}_{j_1 j_2} = \operatorname{argmax}_{\theta_{j_1 j_2}} \ell_{j_1 j_2}(\tilde{\delta}_{j_1}, \tilde{\delta}_{j_2}, \theta_{j_1 j_2}; \mathbf{x}_{j_1}, \mathbf{x}_{j_2}), \tag{5}$$

where  $\theta_{j_1 j_2}$  must satisfy Eq. (2) whose parameters already estimated are fixed by their estimated values, e.g.,  $\theta_{12} = \tilde{\theta}_{12}, \theta_{13} = \tilde{\theta}_{13}, \dots$ , and so on.

After all the bivariate dependence parameters are estimated, the estimates of the trivariate dependence parameters  $\theta_{j_1 j_2 j_3}$  are given by

$$\tilde{\theta}_{j_1 j_2 j_3} = \operatorname{argmax}_{\theta_{j_1 j_2 j_3}} \ell_{j_1 j_2 j_3}(\tilde{\delta}_{j_1}, \tilde{\delta}_{j_2}, \tilde{\delta}_{j_3}, \tilde{\theta}_{j_1 j_2}, \tilde{\theta}_{j_1 j_3}, \tilde{\theta}_{j_2 j_3}, \theta_{j_1 j_2 j_3}; \mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \mathbf{x}_{j_3}),$$

where  $\theta_{j_1 j_2 j_3}$  must satisfy Eq. (2) whose parameters already estimated are fixed by their estimated values, e.g.,  $\theta_{12} = \tilde{\theta}_{12}, \theta_{13} = \tilde{\theta}_{13}, \dots, \theta_{123} = \tilde{\theta}_{123}, \theta_{124} = \tilde{\theta}_{124}, \dots$ , and so on.

In the same manner, we estimate all dependence parameters of higher dimensions in order. Finally,  $\theta_{12\dots d}$  is estimated in a bottom-up fashion. Thus, we can estimate all dependence parameters by repeatedly maximizing functions with one unknown parameter (one equation per dependence parameter). Therefore, IFM can be used to estimate the dependence parameters of the  $d$ -variate FGM distribution no matter how large  $d$  is (even if MLE is infeasible).

If MLE is feasible, an estimator of IFM can be considered as a good starting point for the numerical maximization of the full log-likelihood function (Joe 2014). In Sec. 5.2, the detailed procedure of IFM for the  $d$ -variate FGM distribution is demonstrated in the case of  $d = 3$ .

The proposed algorithm for the  $d$ -variate FGM distribution has the following characteristic remarks that are not seen when IFM is applied to other copulas.

**Remark 1** For a given  $k$ , IFM for the  $d$ -variate FGM distribution requires linear optimization to determine the allowable range of a dependence parameter as  $a_{j_1 \dots j_k} \leq \theta_{j_1 \dots j_k} \leq b_{j_1 \dots j_k}$  where  $a_{j_1 \dots j_k} = \min \theta_{j_1 \dots j_k}$  and  $b_{j_1 \dots j_k} = \max \theta_{j_1 \dots j_k}$  under the condition of Eq. (2), respectively.

For example, let us consider the case that  $d = 3$  and the dependence parameters are estimated in the order of  $\theta_{12} \rightarrow \theta_{23} \rightarrow \theta_{13} \rightarrow \theta_{123}$ . In this case, we have  $a_{12} = \min \theta_{12}$  and  $b_{12} = \max \theta_{12}$  s.t. Eq. (3), respectively. Then,  $a_{12} = -1$  and  $b_{12} = 1$  hold as solutions of these two optimization problems. Eq. (50 can be represented as

$$\tilde{\theta}_{12} = \operatorname{argmax}_{-1 \leq \theta_{12} \leq 1} \ell_{12}(\tilde{\delta}_1, \tilde{\delta}_2, \theta_{12}; \mathbf{x}_1, \mathbf{x}_2).$$

After  $\tilde{\theta}_{12}$  is obtained,  $a_{13}$  and  $b_{13}$  are given by  $\min \theta_{13}$  and  $\max \theta_{13}$  s.t. Eq. (3) where  $\theta_{12} = \tilde{\theta}_{12}$ , respectively. The same discussion can be done for  $a_{23}, b_{23}, a_{123}$ , and  $b_{123}$ . In this way, the parameters of Eq. (2) are fixed by their estimated values one by one.

**Remark 2** Since  $\ell_{j_1 \dots j_k}$  is a single nonlinear equation, some numerical optimization is required to solve  $\operatorname{argmin}_{a_{j_1 \dots j_k} \leq \theta_{j_1 \dots j_k} \leq b_{j_1 \dots j_k}} \ell_{j_1 \dots j_k}$  and  $\operatorname{argmax}_{a_{j_1 \dots j_k} \leq \theta_{j_1 \dots j_k} \leq b_{j_1 \dots j_k}} \ell_{j_1 \dots j_k}$ . A possible initial value of numerical optimization is  $(a_{j_1 \dots j_k} + b_{j_1 \dots j_k})/2$ .

One can use an interior point method as a numerical optimization method for such constrained nonlinear optimization problems (e.g., Mathematica supports an interior point method by the functions *FindMinimum* and *FindMaximum*).

**Remark 3** For a given  $k$ , the order of estimating the  $\binom{d}{k}$  dependence parameters (i.e.,  $\theta_{j_1 \dots j_k}$ 's) is exchangeable. The number of possible ways to estimate  $\theta_{j_1 \dots j_k}$ 's is  $\binom{d}{k}!$ .

For example, when  $d = 3$  and  $k = 2$ , we have  $\theta_{12}, \theta_{13}$ , and  $\theta_{23}$ . The estimation order of these parameters is exchangeable because  $\ell_{12}, \ell_{13}$ , and  $\ell_{23}$  do not depend on  $(\theta_{13}, \theta_{23}), (\theta_{12}, \theta_{23}),$  and  $(\theta_{12}, \theta_{13})$ , respectively. Thus, there are 6 ways to estimate these three parameters in the permutations of  $(\theta_{12}, \theta_{13}, \theta_{23}), (\theta_{12}, \theta_{23}, \theta_{13}), (\theta_{13}, \theta_{12}, \theta_{23}), (\theta_{13}, \theta_{23}, \theta_{12}), (\theta_{23}, \theta_{12}, \theta_{13}),$  and  $(\theta_{23}, \theta_{13}, \theta_{12})$ . In total, we have  $\prod_{k=2}^d \binom{d}{k}!$  ways to estimate all  $2^d - d - 1$  dependence parameters.

The value of  $\tilde{\theta}$  may change depending on the estimation order of  $\theta_{j_1 \dots j_k}$ 's for a small sample. This is because  $a_{j_1 \dots j_k}$  and  $b_{j_1 \dots j_k}$  depend on the values of the dependence parameters already estimated. The estimation using a small sample may yield large estimation errors. As a result, large estimation errors of the dependence parameters obtained at the beginning of the algorithm (e.g.,  $\tilde{\theta}_{12}, \tilde{\theta}_{13}, \dots$ ) tend to shorten allowable ranges  $[a_{j_1 \dots j_k}, b_{j_1 \dots j_k}]$  of other dependence parameters to be estimated. Therefore, we should estimate  $\theta$  by using various ways as long as the computation time is acceptable. If we obtain  $M$  different estimates  $\tilde{\theta}_1, \dots, \tilde{\theta}_M$  by  $M$  ways, we can select the best estimate so that the estimate yields the largest log-likelihood value as  $\operatorname{argmax}_m \ell(\tilde{\delta}, \tilde{\theta}_m; \mathbf{x}_1, \dots, \mathbf{x}_n)$  for  $m = 1, \dots, M$ .

### 4 Asymptotic normality

An estimator of MLE satisfies asymptotic normality under regularity conditions (Cramer 1945). It is also known that an estimator of IFM supports asymptotic normality if the copula satisfies regularity conditions (Joe 2014). However, no studies found that the same property holds in the multivariate FGM distribution.



In this section, therefore, we argue that asymptotic normality of MLE and IFM hold for the FGM distribution. We then compare their asymptotic efficiencies for specific parameter cases.

Without loss of generality, let us consider asymptotic normality of each estimator of MLE and IFM for  $d = 3$  and  $\delta_j = \delta_j$ . Let  $\boldsymbol{\eta}$  be  $(\boldsymbol{\delta}, \boldsymbol{\theta})$  (i.e.,  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_7) = (\delta_1, \delta_2, \delta_3, \theta_{12}, \theta_{13}, \theta_{23}, \theta_{123})$ ), and a parameter with the subscript 0 denotes the true value of the parameter (e.g.,  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\delta}_0$ ). Then, for a random vector of the FGM distribution  $\mathbf{X} = (X_1, X_2, X_3)$ , the following theorems hold under suitable regularity conditions including interchange of differentiation and integration (see pp. 462–463 of Lehmann and Casella (1998) for the consistency and asymptotic normality of MLE and Joe (2014) for the reference of IFM).

**Theorem 1** *The estimator of MLE  $\hat{\boldsymbol{\eta}}$  satisfies the following equation for  $n \rightarrow \infty$ .*

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{D} N(\mathbf{0}, I^{-1}),$$

where  $\xrightarrow{D}$  denotes the convergence in distribution, and  $I$  is the Fisher information matrix, which is given by

$$I = -E \left[ \frac{\partial^2 \log h_3(X_1, X_2, X_3; \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \Bigg|_{\boldsymbol{\eta} = \boldsymbol{\eta}_0} \right], \tag{6}$$

where  $\mathbf{T}$  represents transpose of a vector or matrix, and

$$\begin{aligned} h_3(X_1, X_2, X_3; \boldsymbol{\eta}) = & \prod_{j=1}^3 f_j(X_j; \delta_j) \left( 1 + \theta_{12}(1 - 2F_1(X_1; \delta_1))(1 - 2F_2(X_2; \delta_2)) \right. \\ & + \theta_{13}(1 - 2F_1(X_1; \delta_1))(1 - 2F_3(X_3; \delta_3)) \\ & + \theta_{23}(1 - 2F_2(X_2; \delta_2))(1 - 2F_3(X_3; \delta_3)) \\ & \left. + \theta_{123}(1 - 2F_1(X_1; \delta_1))(1 - 2F_2(X_2; \delta_2))(1 - 2F_3(X_3; \delta_3)) \right). \end{aligned}$$

**Theorem 2** *The estimator of IFM  $\tilde{\boldsymbol{\eta}}$  satisfies the following equation for  $n \rightarrow \infty$ .*

$$\sqrt{n}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{D} N(\mathbf{0}, V),$$

where

$$\begin{aligned} V &= D^{-1} M (D^{-1})^T, \\ D &= E \left[ \frac{\partial s(\mathbf{X}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \Bigg|_{\boldsymbol{\eta} = \boldsymbol{\eta}_0} \right], \end{aligned} \tag{7}$$

$$M = E[s(\mathbf{X}; \boldsymbol{\eta}_0) s(\mathbf{X}; \boldsymbol{\eta}_0)^T], \tag{8}$$

$$s(\mathbf{X}; \boldsymbol{\eta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \delta_1} \log f_1(X_1; \delta_1) \\ \frac{\partial}{\partial \delta_2} \log f_2(X_2; \delta_2) \\ \frac{\partial}{\partial \delta_3} \log f_3(X_3; \delta_3) \\ \frac{\partial}{\partial \theta_{12}} \log h_2(X_1, X_2; \delta_1, \delta_2, \theta_{12}) \\ \frac{\partial}{\partial \theta_{13}} \log h_2(X_1, X_3; \delta_1, \delta_3, \theta_{13}) \\ \frac{\partial}{\partial \theta_{23}} \log h_2(X_2, X_3; \delta_2, \delta_3, \theta_{23}) \\ \frac{\partial}{\partial \theta_{123}} \log h_3(X_1, X_2, X_3; \boldsymbol{\eta}) \end{bmatrix}, \tag{9}$$

$$h_2(X_1, X_2; \delta_1, \delta_2, \theta_{12}) = f_1(X_1; \delta_1) f_2(X_2; \delta_2) (1 + \theta_{12}(1 - 2F_1(X_1; \delta_1))(1 - 2F_2(X_2; \delta_2))).$$

Note that we cannot obtain the explicit forms of  $D$  and  $V$  but can compute them for a specific  $\boldsymbol{\eta}_0$ . Proofs for Theorems 1 and 2 are given in Appendix.

A natural question arises: how much is  $\hat{\boldsymbol{\eta}}$  relatively efficient compared with  $\tilde{\boldsymbol{\eta}}$ ? To take into account this question, we investigate numerical examples. To simplify the asymptotic covariance matrices, let us consider the case in which  $\boldsymbol{\delta}$  is given. That is, the parameters of interest are only  $\boldsymbol{\theta}$ . Suppose  $(\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}) = (0.3, 0.3, 0.3, 0.3)$ . Then, each asymptotic covariance matrix is given by

$$I^{-1} = \begin{pmatrix} 8.48 & 0.721 & 0.721 & -0.486 \\ 0.721 & 8.48 & 0.721 & -0.486 \\ 0.721 & 0.721 & 8.48 & -0.486 \\ -0.486 & -0.486 & -0.486 & 24.32 \end{pmatrix}, \tag{10}$$

$$V = \begin{pmatrix} 8.70 & 0.736 & 0.736 & -0.498 \\ 0.736 & 8.70 & 0.736 & -0.498 \\ 0.736 & 0.736 & 8.70 & -0.498 \\ -0.498 & -0.498 & -0.498 & 24.32 \end{pmatrix}. \tag{11}$$

Note that we use the closed Newton-Cotes formula of degree 3 to numerically compute integrations for  $I^{-1}$  and  $V$ . Since the  $i$ th diagonal element of Eq. (10) is less than that of Eq. (11) for  $i = 1, \dots, 4$ , we can find that MLE is more effective than IFM. For example, the variance of  $\hat{\theta}_{12}$  is around 97% ( $= 8.48/8.70$ ) of the variance of  $\tilde{\theta}_{12}$ .

As another example, let us consider the case of  $(\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}) = (0, 0, 0, 0)$  (i.e., the independent case). Then, each asymptotic covariance matrix is analytically obtained as

$$I^{-1} = V = \begin{pmatrix} 9 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 27 \end{pmatrix}.$$

Hence, the variances of  $\hat{\boldsymbol{\theta}}$  are 100% of those of  $\tilde{\boldsymbol{\theta}}$  if  $(\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}) = (0, 0, 0, 0)$ .

## 5 Simulation study

In this section, we conduct simulation studies to evaluate the effectiveness of our proposed method in the estimation of the multivariate FGM distribution. We first explain how to generate simulation samples from the multivariate FGM distribution. Then, we compare MLE and IFM from the viewpoint of estimation accuracy and computation time and provide some discussion.

### 5.1 Generating the simulation samples

The conditional method (Xu 1996; Joe 2014) allows us to generate samples from the multivariate FGM distribution with arbitrary marginal distributions. Let  $(X_1, \dots, X_d)$  be a random vector of the  $d$ -variate FGM distribution. For  $j = 2, 3, \dots, d$ , we define the conditional distribution functions as follows.

$$C_{j|12\dots j-1}(u_j|u_1, \dots, u_{j-1}) = \Pr[U_j \leq u_j | U_1 = u_1, \dots, U_{j-1} = u_{j-1}].$$

Let  $v_1, \dots, v_d$  be samples from the i.i.d. uniform distribution in the interval  $[0, 1]$ . We construct a random sequence through the following processes.

- Let  $u_1 = C_1^{-1}(v_1)$ .
- Let  $u_2 = C_{2|1}^{-1}(v_2|u_1), \dots,$
- Let  $u_d = C_{d|1\dots d-1}^{-1}(v_d|u_1, \dots, u_{d-1})$ .
- Return  $(x_1, x_2, \dots, x_d) = (F_1^{-1}(u_1; \delta_1), F_2^{-1}(u_2; \delta_2), \dots, F_d^{-1}(u_d; \delta_d))$ .

Consequently, the random sequence  $(x_1, x_2, \dots, x_d)$  follows the  $d$ -variate FGM distribution.

We need to know the closed form of  $C_{j|1\dots j-1}^{-1}(\cdot| \cdot)$  to use the conditional method for the  $d$ -variate FGM distribution. The conditional distribution function is generally given by

$$C_{j|1\dots j-1}(u_j|u_1, \dots, u_{j-1}) = \int_0^{u_j} \frac{c_j(u_1, \dots, u_{j-1}, u)}{c_{j-1}(u_1, \dots, u_{j-1})} du.$$

Since

$$c_j(u_1, \dots, u_j) = c_{j-1}(u_1, \dots, u_{j-1}) + \sum_{p=2}^j \sum_{1 \leq j_1 < \dots < j_p = j} \theta_{j_1 \dots j_p} (1 - 2u_{j_1}) \dots (1 - 2u_{j_p}),$$

it follows that

$$\frac{c_j(u_1, \dots, u_j)}{c_{j-1}(u_1, \dots, u_{j-1})} = 1 + \frac{\sum_{p=2}^j \sum_{1 \leq j_1 < \dots < j_p = j} \theta_{j_1 \dots j_p} (1 - 2u_{j_1}) \dots (1 - 2u_{j_p})}{c_{j-1}(u_1, \dots, u_{j-1})}. \tag{12}$$

By integrating both sides of Eq. (12), we obtain

$$C_{j|1\dots j-1}(u_j|u_1, \dots, u_{j-1}) = (1 + D)u_j - Du_j^2,$$

where

$$D = \frac{\sum_{p=2}^j \sum_{1 \leq j_1 < \dots < j_p = j} \theta_{j_1 \dots j_p} (1 - 2u_{j_1}) \dots (1 - 2u_{j_{p-1}})}{c_{j-1}(u_1, \dots, u_{j-1})}.$$

From  $Du_j^2 - (1 + D)u_j - C_{j|1\dots j-1}(u_j|u_1, \dots, u_{j-1}) = 0$ , the inverse function  $C_{j|1\dots j-1}^{-1}(u_j|u_1, \dots, u_{j-1})$  takes one of the values of

$$\frac{1 + D \pm \sqrt{(1 + D)^2 - 4Du_j}}{2D}, \tag{13}$$

for which it is positive and less than 1. For example, for  $j = 2$ , we have

$$C_{2|1}^{-1}(v_2|u_1) = \frac{1 + \theta_{12}(1 - 2u_1) + \sqrt{(1 + \theta_{12}(1 - 2u_1))^2 - 4\theta_{12}(1 - 2u_1)v_2}}{2\theta_{12}(1 - 2u_1)}.$$

As far as we have found, the signs of Eq. (13) for  $j = 2, 3, 4$  and  $5$  are  $+, -, -, \text{ and } -$ , respectively. Consequently, we can generate samples of the multivariate FGM distribution by using the conditional method with Eq. (13).

### 5.2 Short example

By using simulation data, we demonstrate the procedure of IFM for the  $d$ -variate FGM distribution. The settings of this example are as follows.

1. The sample size is 1000.
2. All the marginal distributions are the i.i.d. uniform distribution in the interval  $[0, 1]$  (i.e., the target parameters are the only dependence parameters).
3. For  $d = 3$ ,  $(\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}) = (0.3, 0.3, 0.3, 0.3)$ .

We generated random samples under the settings and estimate the parameters from the samples.

We estimated the dependence parameters through the following steps.

Step 1:  $\tilde{\theta}_{12} = \arg \max_{\theta_{12}} \ell_{12} = \arg \max_{-1 \leq \theta_{12} \leq 1} \ell_{12} = 0.298.$

Step 2:  $\tilde{\theta}_{13} = \arg \max_{\theta_{13}, \theta_{12}=0.298} \ell_{13} = \arg \max_{-1 \leq \theta_{13} \leq 1} \ell_{13} = 0.272.$

Step 3:  $\tilde{\theta}_{23} = \arg \max_{\theta_{23}, \theta_{12}=0.298, \theta_{13}=0.272} \ell_{23} = \arg \max_{-0.431 \leq \theta_{23} \leq 0.974} \ell_{23} = 0.365.$

Step 4:  $\tilde{\theta}_{123} = \arg \max_{\theta_{123}, \theta_{12}=0.298, \theta_{13}=0.272, \theta_{23}=0.365} \ell = \arg \max_{-0.609 \leq \theta_{123} \leq 0.609} \ell = 0.222.$

Note that the allowable ranges of the dependence parameters are calculated based on Remark 1 in each step.

As mentioned in Remark 3, Step 1, 2, and 3 are exchangeable although we estimated the dependence parameters in the order of  $\theta_{12}, \theta_{13}$ , and  $\theta_{23}$  in this example. If the order is changed, the estimation result might be different from the above result. To illustrate this phenomenon, we conducted another experiment by changing the sample size from 1000 to 100. We estimated all dependence parameters by MLE and IFM with possible six estimation orders (see Table 1). Note that  $(\alpha, \beta)$  below each estimated value in Table 1 represents the allowable range of the corresponding dependence parameter (i.e.,  $\alpha = a_{j_1 \dots j_k}$  and  $\beta = b_{j_1 \dots j_k}$ ). For example,  $\theta_{12}$  was estimated in the interval  $[-1, 1]$  in the estimation order of  $\theta_{12} \rightarrow \theta_{13} \rightarrow \theta_{23} \rightarrow \theta_{123}$ .

As shown in Table 1, there were 3 different types of estimation results for IFM, and all  $\tilde{\theta}_{123}$  were 0 under the conditions of Eq. (2) and  $(\theta_{12}, \theta_{13}, \theta_{23}) = (\tilde{\theta}_{12}, \tilde{\theta}_{13}, \tilde{\theta}_{23})$ . The reason why we obtained such results due to the difference of the estimation order mainly depended on the accumulation of estimation errors in the process of IFM. For example, the estimated values  $\tilde{\theta}_{13}$  were different between the cases that the dependence parameters were estimated in the orders of  $\theta_{12} \rightarrow \theta_{13} \rightarrow \theta_{23} \rightarrow \theta_{123}$  and  $\theta_{12} \rightarrow \theta_{23} \rightarrow \theta_{13} \rightarrow \theta_{123}$  (i.e., 0.707 and 0.518). While the allowable range of  $\theta_{13}$  thirdly estimated in the latter case was  $[-0.156, 0.518]$ , that secondly estimated in the former case was  $[-1, 1]$ .  $\tilde{\theta}_{13}$  could not be 0.707 in the latter case. Therefore, the value of  $\tilde{\theta}$  changed depending on the estimation order.

The result of MLE was also different from the results of IFM. Based on the log-likelihood values, the result of MLE should be chosen as the best estimate in this example.

**Table 1** Comparison among simulation results of estimating  $(\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}) = (0.3, 0.3, 0.3, 0.3)$  by possible 6 IFM ways and MLE

Method (Estimation order)	$\theta_{12}$	$\theta_{13}$	$\theta_{23}$	$\theta_{123}$	Log likelihood
IFM ( $\theta_{12} \rightarrow \theta_{13} \rightarrow \theta_{23} \rightarrow \theta_{123}$ )	0.181 (-1, 1)	0.707 (-1, 1)	0.474 (-0.112, 0.474)	0 (0, 0)	4.871
IFM ( $\theta_{12} \rightarrow \theta_{23} \rightarrow \theta_{13} \rightarrow \theta_{123}$ )	0.181 (-1, 1)	0.518 (-0.156, 0.518)	0.663 (-1, 1)	0 (0, 0)	4.898
IFM ( $\theta_{13} \rightarrow \theta_{12} \rightarrow \theta_{23} \rightarrow \theta_{123}$ )	0.181 (-1, 1)	0.707 (-1, 1)	0.474 (-0.112, 0.474)	0 (0, 0)	4.871
IFM ( $\theta_{13} \rightarrow \theta_{23} \rightarrow \theta_{12} \rightarrow \theta_{123}$ )	0.370 (0.370, 0.956)	0.707 (-1, 1)	0.663 (-1, 1)	0 (0, 0)	5.004
IFM ( $\theta_{23} \rightarrow \theta_{12} \rightarrow \theta_{13} \rightarrow \theta_{123}$ )	0.181 (-1, 1)	0.518 (-0.156, 0.518)	0.663 (-1, 1)	0 (0, 0)	4.898
IFM ( $\theta_{23} \rightarrow \theta_{13} \rightarrow \theta_{12} \rightarrow \theta_{123}$ )	0.370 (0.370, 0.956)	0.707 (-1, 1)	0.663 (-1, 1)	0 (0, 0)	5.004
MLE	0.298	0.639	0.610	0.048	5.065

### 5.3 Estimation accuracy

In this subsection, we investigated the estimation accuracy of the proposed algorithm through Monte Carlo simulation under the conditions that parameters of marginal distributions  $\delta$  are known and unknown.

#### 5.3.1 When $\delta$ is known

We again considered that all the marginal distributions were given, and  $\delta$  was known. For  $d = 4$ , we conducted the simulation under the following settings.

1. The sample sizes  $n$  were 100, 1000, and 10,000, and the number of iteration times  $m$  was 100.
2. All the marginal distributions were the i.i.d. uniform distribution in the interval  $[0, 1]$ . This means that the target parameters are only the dependence parameters.
3. We considered the following four situations of the dependence parameters.
  - (a)  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234})$   
 $= (0, 0, 0, 0, 0, 0, 0.2, 0.2, 0.2, 0.2, 0.2).$
  - (b)  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234})$   
 $= (0, 0, 0, 0, 0, 0, 0.2, 0.2, 0.2, 0.2, 0.5).$
  - (c)  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234})$   
 $= (0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2).$
  - (d)  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234})$   
 $= (0, 0, 0, 0, 0, 0, -0.2, -0.2, -0.2, -0.2, -0.2).$

Under the settings, we calculated the following Bias and root-mean-squared error (RMSE) as indicators of estimation accuracy for each estimator of IFM.

$$Bias(\tilde{\theta}) = \begin{bmatrix} \frac{1}{m} \sum_{i=1}^m (\tilde{\theta}_{12,i} - \theta_{12(0)}) \\ \vdots \\ \frac{1}{m} \sum_{i=1}^m (\tilde{\theta}_{1234,i} - \theta_{1234(0)}) \end{bmatrix},$$

$$RMSE(\tilde{\theta}) = \begin{bmatrix} \sqrt{\frac{1}{m} \sum_{i=1}^m (\tilde{\theta}_{12,i} - \theta_{12(0)})^2} \\ \vdots \\ \sqrt{\frac{1}{m} \sum_{i=1}^m (\tilde{\theta}_{1234,i} - \theta_{1234(0)})^2} \end{bmatrix},$$

where  $\tilde{\theta}_{\cdot,i}$  is  $\tilde{\theta}$ 's estimate obtained in the  $i$ th iteration step and  $\tilde{\theta}_{\cdot(0)}$  is the true value of  $\tilde{\theta}$ . Bias and RMSE represent the average and dispersion of the estimation errors, respectively.

The numerical results are presented in Tables 2, 3, 4, and 5. Note that MLE cannot work in the case of  $n = 10,000$  due to too much computation time for 100 iterations (represented as N/A). Besides, only one estimation order of  $\tilde{\theta}$  was considered

**Table 2** Simulation results of estimating  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234}) = (0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2)$  based on the average of 100 iterations

Bias	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$
$n = 100$	IFM	-0.083	0.043	0.033	-0.036	0.048	0.010	0.105	0.110	0.109	0.169	0.194
	MLE	-0.082	0.027	0.021	-0.029	0.037	0.009	0.106	0.081	0.048	0.080	0.137
$n = 1000$	IFM	0.007	0.007	-0.013	-0.020	0.022	-0.006	0.020	0.031	0.017	0.007	0.086
	MLE	0.004	0.005	-0.013	-0.021	0.021	-0.007	0.036	0.050	0.015	-0.004	0.035
$n = 10,000$	IFM	0.005	-0.002	0.001	-0.004	0.001	0.001	0.001	0.002	0.001	0.001	0.003
	MLE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RMSE	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$
$n = 100$	IFM	0.296	0.268	0.305	0.276	0.295	0.307	0.316	0.310	0.292	0.284	0.334
	MLE	0.273	0.225	0.253	0.249	0.281	0.290	0.313	0.332	0.334	0.317	0.321
$n = 1000$	IFM	0.092	0.087	0.090	0.094	0.096	0.093	0.168	0.152	0.161	0.159	0.205
	MLE	0.090	0.084	0.088	0.093	0.094	0.091	0.158	0.147	0.159	0.162	0.219
$n = 10,000$	IFM	0.032	0.029	0.029	0.029	0.033	0.029	0.057	0.048	0.053	0.050	0.079
	MLE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A





**Table 4** Simulation results of estimating  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{234}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{1234}) = (0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)$  based on the average of 100 iterations

Bias	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$
$n = 100$	IFM	- 0.078	0.035	0.024	- 0.033	0.044	0.039	0.112	0.117	0.073	0.096	0.098
	MLE	- 0.050	0.047	0.048	- 0.004	0.046	0.035	0.094	0.059	0.020	0.056	0.098
$n = 1000$	IFM	0.006	0.007	- 0.012	- 0.02	0.02	- 0.007	0.018	0.036	0.013	- 0.002	0.084
	MLE	0.006	0.007	- 0.007	- 0.017	0.023	- 0.006	0.023	0.048	0.019	0.000	0.040
$n = 10,000$	IFM	0.005	- 0.002	0.001	- 0.003	0.001	0.000	0.000	0.003	0.001	0.002	0.001
	MLE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RMSE	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$
$n = 100$	IFM	0.294	0.264	0.305	0.271	0.294	0.299	0.290	0.283	0.287	0.249	0.297
	MLE	0.266	0.223	0.257	0.243	0.274	0.286	0.297	0.322	0.343	0.310	0.280
$n = 1000$	IFM	0.090	0.088	0.089	0.097	0.094	0.095	0.165	0.152	0.156	0.164	0.208
	MLE	0.086	0.084	0.084	0.094	0.093	0.093	0.155	0.149	0.158	0.162	0.229
$n = 10,000$	IFM	0.032	0.028	0.029	0.029	0.032	0.030	0.055	0.048	0.051	0.050	0.081
	MLE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

**Table 5** Simulation results of estimating  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234}) = (0, 0, 0, 0, 0, -0.2, -0.2, -0.2, -0.2, -0.2, -0.2)$  based on the average of 100 iterations

Bias	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$
$n = 100$	IFM	-0.083	0.041	0.033	-0.040	0.038	0.016	-0.079	-0.080	-0.202	-0.187	-0.174
	MLE	-0.080	0.027	0.031	-0.031	0.036	0.027	-0.071	-0.095	-0.118	-0.114	-0.159
$n = 1000$	IFM	0.007	0.009	-0.012	-0.020	0.020	-0.009	0.016	0.029	-0.02	-0.061	-0.146
	MLE	-0.002	0.002	-0.018	-0.026	0.014	-0.015	-0.006	0.010	-0.019	-0.041	-0.063
$n = 10,000$	IFM	0.005	-0.002	0.001	-0.004	0.001	0.000	0.001	0.003	0.002	0.002	0.055
	MLE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RMSE	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$
$n = 100$	IFM	0.296	0.275	0.309	0.276	0.287	0.312	0.291	0.304	0.320	0.316	0.327
	MLE	0.274	0.231	0.269	0.252	0.259	0.295	0.310	0.360	0.346	0.338	0.356
$n = 1000$	IFM	0.092	0.088	0.091	0.094	0.093	0.095	0.165	0.154	0.153	0.159	0.254
	MLE	0.090	0.084	0.087	0.094	0.093	0.095	0.153	0.136	0.152	0.165	0.243
$n = 10,000$	IFM	0.032	0.028	0.029	0.029	0.033	0.029	0.057	0.047	0.053	0.051	0.101
	MLE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

to finish the experiments within reasonable computation time (e.g., there are 17280 estimation orders for  $d = 4$  as discussed in Remark 3).

The estimation accuracy improved with a power of 10 samples. To obtain estimates with Bias of less than 0.1, we require approximately 100, 1000, and 10,000 samples for  $\theta_{j_1j_2}$ ,  $\theta_{j_1j_2j_3}$ , and  $\theta_{1234}$ , respectively. As for RMSE, ratio values of the results of IFM to those of MLE for the corresponding dependence parameters are close to 1 in most of the cases.

### 5.3.2 When $\delta$ is unknown

We next investigated the estimation accuracy when the marginal distributions were given, but  $\delta$  was unknown. In this case, it is expected that the accuracy of IFM worsen because the estimation accuracy for the dependence parameters is affected by the error for the marginal parameters. To check how much the performance worsens, we conducted the simulation under the following settings for  $d = 4$ .

1. The sample size  $n$  was 100, and the number of iteration times  $m$  was 100.
2. All the marginal distributions were an i.i.d. exponential distribution. We fixed the marginal parameters as  $(\delta_1, \delta_2, \delta_3, \delta_4) = (1, 1, 1, 1)$ .
3. The dependence parameters were fixed as  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234}) = (0, 0, 0, 0, 0, 0, 0.2, 0.2, 0.2, 0.2, 0.2)$ .

The estimation results of the marginal parameters and dependence parameters are summarized in Tables 6 and 7, respectively. In Table 6, since  $n = 100$  was enough to estimate the marginal parameters precisely, all biases are close to 0. Hence, the estimation accuracy for the dependence parameters did not significantly worsen even if the marginal parameters were unknown. Bias and RMSE for the dependence parameters in Table 7 were almost the same as those for the dependence parameters shown in the result of  $n = 100$  of Table 2. Therefore, we infer from these results that IFM and MLE for the parameter estimation of the FGM distribution work well under the condition that the sample size is enough to estimate  $\delta$  at least.

**Table 6** Simulation results of estimating  $(\delta_1, \delta_2, \delta_3, \delta_4) = (1, 1, 1, 1)$  based on the average of 100 iterations

	Method	$\delta_1$	$\delta_2$	$\delta_3$	$\delta_4$
Bias	IFM	- 0.016	- 0.020	0.000	- 0.018
	MLE	- 0.018	- 0.024	- 0.001	- 0.190
RMSE	IFM	0.092	0.113	0.095	0.102
	MLE	0.093	0.116	0.096	0.103

**Table 7** Simulation results of estimating  $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}, \theta_{34}, \theta_{123}, \theta_{124}, \theta_{134}, \theta_{234}, \theta_{1234}) = (0, 0, 0, 0, 0, 0, 0.2, 0.2, 0.2, 0.2, 0.2)$  if  $\delta$  is unknown based on the average of 100 iterations

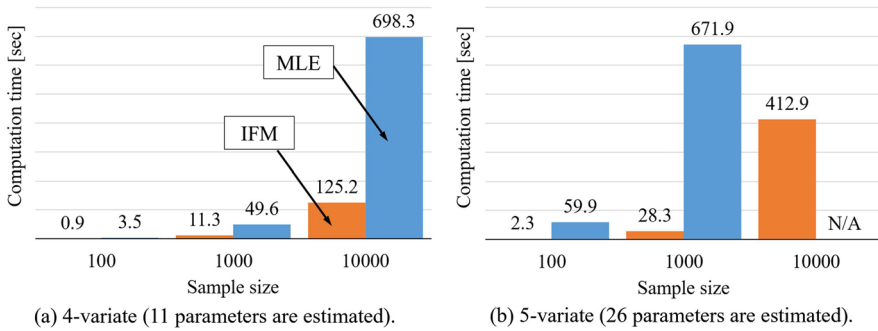
Bias	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$
	IFM	- 0.082	0.046	0.036	- 0.040	0.055	0.010
	MLE	- 0.081	0.022	0.025	- 0.031	0.041	0.006
	Method	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$	
	IFM	0.102	0.110	0.109	0.171	0.201	
	MLE	0.103	0.084	0.046	0.087	0.153	
RMSE	Method	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{23}$	$\theta_{24}$	$\theta_{34}$
	IFM	0.297	0.272	0.308	0.283	0.296	0.304
	MLE	0.273	0.227	0.254	0.261	0.283	0.292
	Method	$\theta_{123}$	$\theta_{124}$	$\theta_{134}$	$\theta_{234}$	$\theta_{1234}$	
	IFM	0.315	0.314	0.280	0.291	0.338	
	MLE	0.308	0.342	0.334	0.323	0.3377	

### 5.4 Computation time

In this subsection, we discuss the investigation of the computation times of MLE and IFM. We implemented IFM for the  $d$ -variate FGM distribution in Mathematica 11.3 and a PC equipped with Intel(R)Core(TM) i7-6700k 4.00-GHz, 32GB-RAM running on Windows 10 professional 64bit. We conducted the simulation under the following settings.

1. The dimensions  $d$  were 4 and 5.
2. The sample sizes  $n$  were 100, 1000, and 10,000.
3. All the marginal distributions were the i.i.d. uniform distribution in the interval  $[0, 1]$ . This means that the target parameters are only the dependence parameters.
4. For  $d = 4$ , the values of the dependence parameters are as follows:  $\theta_{j_1j_2} = 0, \theta_{j_1j_2j_3} = 0.2$ , and  $\theta_{1234} = 0.2$  for  $1 \leq j_1 < j_2 < j_3 \leq 4$ .
5. For  $d = 5$ , the values of the dependence parameters are as follows:  $\theta_{j_1j_2} = 0, \theta_{j_1j_2j_3} = 0, \theta_{j_1j_2j_3j_4} = 0.2$ , and  $\theta_{12345} = 0.2$  for  $1 \leq j_1 < j_2 < j_3 < j_4 \leq 5$ .

Figure 1 illustrates the computation times to estimate the dependence parameters under the above situation. We can find that IFM works faster than MLE in all cases, and both MLE and IFM require more computational time as  $d$  and  $n$  increase. If  $d = 5$  and  $n = 10,000$ , MLE cannot work due to running out of memory (represented as N/A). Therefore, to estimate the parameters of the multivariate FGM distribution, we should not use MLE but IFM from the viewpoint of computational cost. A theoretical analysis of the computational complexity of IFM is required for future work.



**Fig. 1** Average computation times of estimating all dependence parameters of  $d$ -variate FGM distribution for  $d = 4, 5$  (number of iterations: 5)

## 6 Case study

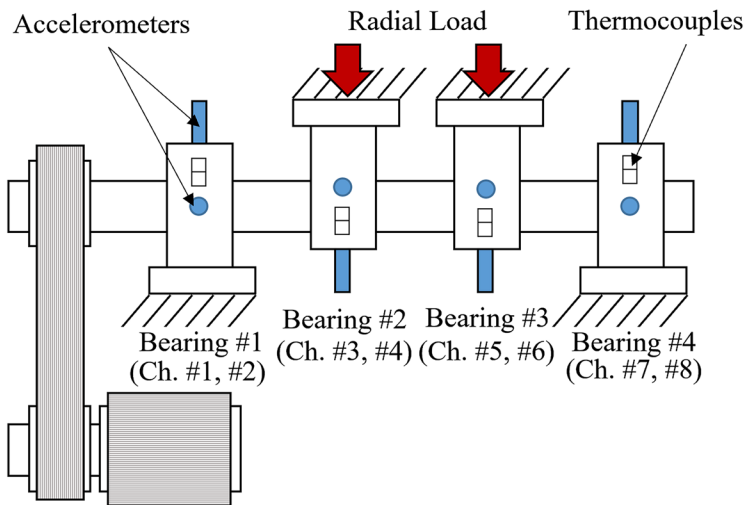
In the previous section, we showed several aspects of the parameter estimation scheme proposed in this paper by utilizing simulation data. In this section, in turn, we depict another numerical example based on the actually collected data set. This is called *Bearing data set*, which is available in NASA Ames Prognostics Data Repository (Lee et al. 2007).

### 6.1 Bearing data set

This data set was collected for the purpose of the investigation of how four bearings with a coaxial rotating shaft degraded in time. For this case study, we employ “Set No. 1” data set from the original data set. This “Set No.1” contains 2156 files, and each file has 20480 data points of vibration signal yielded in one second from 4 working bearings with 2 data-collection channels of accelerometers on the test rig (see Fig. 2). Therefore one data file has 20480 lines, and each line has 8 numerical values (4 bearings  $\times$  2 channels). Also, one file was saved as a log file every 10 minutes (N.B., the first 43 files are logged every 5 minutes). Hence we obtain about 44.1 million ( $\approx 2156 \times 20480$ ) lines of 8-dimension vibration signal values.

In order to reduce the size of the data set, we extracted the first line from every 2156 files. And from the structure of the test rig, we decided to use 4 channels of #1, #3, #5, and #7 from bearing #1, #2, #3, and #4, respectively. After such a data preparation, we finally obtain 2156 actual samples of 4-variate data set as follows ( $i = 1, 2, \dots, 2156$ ).

- $x_{i1}$  = the  $i$ th vibration signal (unit: arbitrary) of channel #1 in bearing #1.
- $x_{i2}$  = the  $i$ th vibration signal (unit: arbitrary) of channel #3 in bearing #2.
- $x_{i3}$  = the  $i$ th vibration signal (unit: arbitrary) of channel #5 in bearing #3.
- $x_{i4}$  = the  $i$ th vibration signal (unit: arbitrary) of channel #7 in bearing #4.



**Fig. 2** Illustration of test rig [cited from Qiu et al. (2006)]. Each bearing has 2 channels for data collection. Bearing #1 has channels #1 and #2, bearing #2 has channels #3 and #4, bearing #3 has channels #5 and #6, and bearing #4 has channels #7 and #8, respectively

Note that the unit of the vibration signal is not indicated in the original references of *Bearing data set* (Qiu et al. 2006; Lee et al. 2007). Since the vibration signal was collected from the accelerometers, we can guess that the unit is any of voltage (mV) [or (V)] and acceleration ( $m/s^2$ ) which is proportional to a voltage value. For this reason, we assume that the vibration signal has an arbitrary unit in this study.

In this paper, we use this data set to find some dependence property among four bearings via their vibration signal data. Since these four bearings are operated by a single rotating shaft, it is expected that some common factors may affect the degradation of the bearings. For this, we apply 4-variate FGM copula to the data set and estimate its 11 dependence parameters.

## 6.2 Finding marginal distribution

We separate the data set into three sets from the viewpoint of the size. We call them Data-A, Data-B, and Data-C, respectively. Data-C is the overall data (2156 lines). To make Data-A, we extracted the first 100 lines from Data-C. This means that this set was collected from the early stage of the operation time of four bearings. The second one, Data-B, consists of 1000 lines (i.e., Data-A plus successive 900 lines) to grasp the dependence property until the middle part of the data set. Data-C is used for the overall time domain analysis.

For each data set, we need to find the well-fitted marginal distribution firstly. Since the normal distribution was not fit these data sets, we chose the generalized  $t$ -distribution with the following probability density function:

$$f(x; \delta) = \frac{[1 + (\frac{x-a}{b\sqrt{c}})^2]^{-\frac{c+1}{2}}}{b\sqrt{c} B[\frac{c}{2}, \frac{1}{2}]}, \tag{14}$$

where  $\delta = (a, b, c)$  and  $B[\alpha, \beta]$  is the beta function defined by

$$\int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt.$$

This distribution has a location parameter  $a$  and a scale parameter  $b$ . The parameter  $c$  denotes the degree of freedom. Therefore if  $a = 0$  and  $b = 1$ , Eq. (14) describes the Student’s  $t$ -distribution.

Table 8 summarizes the estimation results of the marginal  $t$ -distribution and their  $p$  values by employing the Kolmogorov–Smirnov goodness-of-fit test. The right-most column also lists the  $p$  values if the normal distribution is applied. Therefore we choose  $t$ -distribution for the marginal, because all the cases are not rejected at the significance level of 0.05.

### 6.3 Parameter estimation and discussion

We estimated all the parameters included in the 4-variate FGM copula by the proposed method. Note that only one estimation order of  $\tilde{\theta}$  was considered to finish

**Table 8** List of estimated parameters of marginal  $t$ -distribution. Normal distribution is not suitable because of the several small  $p$  values

Channel 1	$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$p$ ( $t$ -dist)	$p$ (normal dist.)
Data-A	− 0.127	0.101	21.3	0.798	0.820
Data-B	− 0.122	0.0914	42.0	0.402	0.508
Data-C	− 0.119	0.0877	16.0	0.359	0.349
Channel 3	$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$p$ ( $t$ -dist)	$p$ (normal dist.)
Data-A	− 0.113	0.111	73.9	0.667	0.632
Data-B	− 0.116	0.0940	25.5	0.508	0.307
Data-C	− 0.116	0.0905	26.8	0.183	0.128
Channel 5	$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$p$ ( $t$ -dist)	$p$ (normal dist.)
Data-A	− 0.123	0.152	3.95	0.981	0.443
Data-B	− 0.118	0.109	7.96	0.876	0.402
Data-C	− 0.115	0.103	8.28	0.789	0.0157
Channel 7	$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$p$ ( $t$ -dist)	$p$ (normal dist.)
Data-A	− 0.122	0.117	6.82	0.843	0.415
Data-B	− 0.114	0.0905	4.63	0.863	$2.88 \times 10^{-3}$
Data-C	− 0.116	0.0743	4.73	0.558	$1.11 \times 10^{-5}$

**Table 9** List of estimated parameters by IFM. Subscript numbers 1, 2, 3, and 4 of the  $\theta$  correspond to the channel numbers 1, 3, 5, and 7, respectively

Data	$\tilde{\theta}_{12}$	$\tilde{\theta}_{13}$	$\tilde{\theta}_{14}$	$\tilde{\theta}_{23}$	$\tilde{\theta}_{24}$	$\tilde{\theta}_{34}$	$\tilde{\theta}_{123}$	$\tilde{\theta}_{124}$	$\tilde{\theta}_{134}$	$\tilde{\theta}_{234}$	$\tilde{\theta}_{1234}$
Data-A	0.308	-0.207	0.0634	-0.535	-0.623	0.561	-0.140	0.00618	0.168	0.0224	0.227
Data-B	0.465	-0.161	-0.0601	-0.385	-0.486	0.261	-0.0434	-0.0280	0.199	0.0461	0.0644
Data-C	0.487	-0.130	0.0179	-0.400	-0.452	0.262	-0.0162	0.0431	0.0810	0.0620	-0.0110

the experiment within acceptable computation time. Table 9 is a list of them for each data set. In the statistical sense, the results of Data-C are the most meaningful because the size of the data set is the largest (2156 data points) among the three. The values in each row vary by the combination of bearings. Actually, it is not easy to infer the actual dependence structure among the bearings on the test rig from such information seen in this table, but we dare to state the degree of dependence by categorizing the estimated values. That is, we set three categories of the dependence as *positive*, *negative*, and *almost zero* between two bearings. By following this categorization,  $\tilde{\theta}_{12}$ ,  $\tilde{\theta}_{34}$  are judged as *positive*,  $\tilde{\theta}_{13}$ ,  $\tilde{\theta}_{23}$ , and  $\tilde{\theta}_{24}$  are *negative*, and  $\tilde{\theta}_{14}$  is *almost zero*.

Translating the subscript numbers 1, 2, 3, and 4 of  $\theta$  into the channel numbers 1, 3, 5, and 7, respectively, we can infer that the bearings #1 and #4 are almost independent because  $\tilde{\theta}_{14}$  belongs to the category of *almost zero*. It is understandable that this result from the illustration of Fig. 2 is possible since the bearing #1 and #4 are placed at both ends on the test rig equipment. It is reasonable that they may yield mutually independent vibration signals.

On the other hand, the degree of the *positive* and *negative* dependence between two bearings seems to also suggest some structural dependence depicted in Fig. 2. In the case of *positive*, the pair of bearings #1 and #2 adjoin, and the direction of the radial load is opposite. The same relationship can be seen at the pair of bearings #3 and #4. Therefore we can presume that these pairs of bearings may fail dependently in the future. Regarding this matter, Qiu et al. (2006) reported that an inner race defect was discovered in the bearing #3 and a roller element defect and outer race defect were found in the bearing #4 at the end of the testing. This consequence results in supporting our inference stated above.

The last category, denoted as *negative*, contains three pairs of bearings as (#1, #3), (#2, #4), and (#2, #3). Among the three, the first two pairs of bearings suffer the opposite radial load direction, but have more distance than the case of *positive*. Whereas in the case of (#2, #3), these bearings take the same direction of the radial load and are equipped closely. It is not clear that how these factors affect that the estimated values tend to be negative. It may be needed to do more research from the viewpoint of mechanical engineering and other theoretical consideration to explain these interesting relationships between the estimated values and the physical layout of the equipment and other factors of the bearings.



Moreover, it becomes more complicated to explain the dependence structure and estimated values among three or more bearings (i.e.,  $\theta_{123}, \dots, \theta_{1234}$ ). It is desired that the reasons and implications for these phenomena are revealed in future studies. In any case, however, we have confirmed that the proposed algorithm for the parameter estimation of the FGM distribution worked well.

## 7 Conclusion

We proposed an algorithm for estimating the parameters of the multivariate FGM distribution in practical computation time. The computational complexity of MLE was reduced by using IFM. Because of this, IFM can estimate all the dependence parameters of the multivariate FGM distribution no matter how high the dimension  $d$  is. We then revealed that the estimators of IFM satisfy asymptotic normality for the multivariate FGM distribution. Therefore, The proposed algorithm is useful for estimating the parameters of the FGM distribution.

From the viewpoint of reliability engineering, the proposed algorithm can be applied to a quantitative evaluation of the dependence among system components. If we have the lifetime data of system components, we can find the latent dependence structure among them by estimating the parameters of the multivariate FGM distribution with arbitrary marginal distributions.

## Appendix: Proofs

Before we prove Theorems 1 and 2, we introduce a lemma and theorem as follows.

**Lemma 1** *Let  $(X_1, X_2, X_3)$  be a random vector of the trivariate FGM distribution. Then,  $E[s(X_1, X_2, X_3; \eta_0)] = \mathbf{0}$  and  $E[\frac{\partial}{\partial \eta} \log h_3(X_1, X_2, X_3; \cdot) |_{\eta=\eta_0}] = \mathbf{0}$  hold.*

**Proof of Lemma 1** (9), we prove that each element of  $E[s(\cdot; \eta_0)]$  is 0, that is,

$$\begin{aligned}
 E\left[\frac{\partial}{\partial \delta_j} \log f_j(X_j; \cdot)\right] &= 0, \\
 E\left[\frac{\partial}{\partial \theta_{jk}} \log h_2(X_j, X_k; \cdot)\right] &= 0, \\
 E\left[\frac{\partial}{\partial \theta_{123}} \log h_3(X_1, X_2, X_3; \cdot)\right] &= 0,
 \end{aligned}$$

for  $1 \leq j < k \leq 3$ . First, we have

$$\begin{aligned}
 E \left[ \frac{\partial}{\partial \delta_j} \log f_j(X_j; \cdot) \right] &= \int_{-\infty}^{\infty} f_j(x_j; \cdot) \frac{\partial}{\partial \delta_j} \log f_j(x_j; \cdot) dx_j \\
 &= \int_{-\infty}^{\infty} \frac{\partial}{\partial \delta_j} f_j(x_j; \cdot) dx_j \\
 &= \frac{\partial}{\partial \delta_j} \int_{-\infty}^{\infty} f_j(x_j; \cdot) dx_j = 0,
 \end{aligned}$$

where we assume that the order of the differentiation and integration can be changed in the third step. In the same manner, we obtain

$$\begin{aligned}
 E \left[ \frac{\partial}{\partial \theta_{jk}} \log h_2(X_j, X_k; \cdot) \right] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{jk}(x_j, x_k; \cdot) \frac{\partial}{\partial \theta_{jk}} \log h_2(x_j, x_k; \cdot) dx_j dx_k \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_{jk}} h_2(x_j, x_k; \cdot) dx_j dx_k \\
 &= \frac{\partial}{\partial \theta_{jk}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(x_j, x_k; \cdot) dx_j dx_k = 0,
 \end{aligned}$$

and

$$\begin{aligned}
 E \left[ \frac{\partial}{\partial \theta_{123}} \log h_3(X_1, X_2, X_3; \cdot) \right] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(x_1, x_2, x_3; \cdot) \frac{\partial}{\partial \theta_{123}} \log h_3(x_1, x_2, x_3; \cdot) dx_1 dx_2 dx_3 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_{123}} h_3(x_1, x_2, x_3; \cdot) dx_1 dx_2 dx_3 \\
 &= \frac{\partial}{\partial \theta_{123}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(x_1, x_2, x_3; \cdot) dx_1 dx_2 dx_3 = 0.
 \end{aligned}$$

Therefore,  $E[s(\cdot; \eta_0)] = \mathbf{0}$ . Also,  $E[\frac{\partial}{\partial \eta} \log h_3(X_1, X_2, X_3; \cdot)] = \mathbf{0}$  holds in the same manner. Hence, the proof is complete. □

**Proof of Theorem 1** We provide a sketch of the proof of Theorem 1. Suppose  $Y, Y_1, \dots, Y_n$  are i.i.d. random vectors of the  $d$ -variate FGM distribution. By applying the mean value theorem for the maximum log likelihood function at  $\eta_0$ , we obtain

$$\begin{aligned}
 \mathbf{0} &= \left. \frac{\partial \mathcal{L}(\eta; Y_1, \dots, Y_n)}{\partial \eta} \right|_{\eta = \hat{\eta}} \\
 &= \left. \frac{\partial \mathcal{L}(\eta; Y_1, \dots, Y_n)}{\partial \eta} \right|_{\eta = \eta_0} + K(\bar{\eta})(\hat{\eta} - \eta_0),
 \end{aligned} \tag{15}$$

where a matrix  $K(\bar{\eta})$  is given by

$$K(\bar{\boldsymbol{\eta}}) = \frac{\partial^2 \ell(\boldsymbol{\eta}; \mathbf{Y}_1, \dots, \mathbf{Y}_n)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \Big|_{\boldsymbol{\eta}=\bar{\boldsymbol{\eta}}},$$

and  $\bar{\boldsymbol{\eta}} \equiv \text{diag}(\alpha_1, \dots, \alpha_7) \hat{\boldsymbol{\eta}} + \text{diag}(1 - \alpha_1, \dots, 1 - \alpha_7) \boldsymbol{\eta}_0$  for  $(\alpha_1, \dots, \alpha_7) \in [0, 1]^7$  (N.B.,  $\text{diag}(a_1, \dots, a_7)$  is a diagonal matrix whose elements starting in the upper left corner are  $a_1, \dots, a_7$ ). Let us assume the invertibility of  $I$ . From Eq. (15), we have

$$\begin{aligned} & \sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \\ &= - \left[ \frac{1}{n} K(\bar{\boldsymbol{\eta}}) \right]^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\eta}; \mathbf{Y}_1, \dots, \mathbf{Y}_n)}{\partial \boldsymbol{\eta}} \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \\ &= - \left[ \frac{1}{n} K(\bar{\boldsymbol{\eta}}) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\eta}} \log h_3(\mathbf{Y}_i; \boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \end{aligned} \tag{16}$$

The first factor on the right-hand side of Eq. (16) can be applied to the law of large numbers because it is divided by the sample size  $n$ . By assuming  $\hat{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}_0$  for  $n \rightarrow \infty$ ,  $\bar{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}_0$  also holds for  $n \rightarrow \infty$ . Recalling Eq. (6), we obtain

$$\frac{1}{n} K(\bar{\boldsymbol{\eta}}) \xrightarrow{p} -E \left[ \frac{\partial^2 \log h_3(\mathbf{Y}; \cdot)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \right] = I. \tag{17}$$

The second factor can be applied to the multivariate central limit theorem because it is divided by  $\sqrt{n}$ . From Lemma 1 and Eq. (17), we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\eta}} \log h_3(\mathbf{Y}_i; \boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \xrightarrow{D} N(\mathbf{0}, I). \tag{18}$$

Therefore, by applying Slutsky’s theorem (see e.g., Cramer (1945), p. 254) to Eqs. (17) and (18), the following equation holds.

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{D} N(\mathbf{0}, I^{-1}).$$

Hence, the proof is complete. □

**Proof of Theorem 2** By the same analogy with the proof of Theorem 1, we provide a sketch of the proof of Theorem 2. Suppose  $\mathbf{Y}, \mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d. random vectors of the  $d$ -variate FGM distribution. Let us define  $\mathbf{g}(\mathbf{Y}_1, \dots, \mathbf{Y}_n; \boldsymbol{\eta})$  as follows.

$$\begin{aligned} & \mathbf{g}(\mathbf{Y}_1, \dots, \mathbf{Y}_n; \boldsymbol{\eta}) \\ &= \left( \frac{\partial \ell_1}{\partial \delta_1}, \frac{\partial \ell_2}{\partial \delta_2}, \frac{\partial \ell_3}{\partial \delta_3}, \frac{\partial \ell_{12}}{\partial \theta_{12}}, \frac{\partial \ell_{13}}{\partial \theta_{13}}, \frac{\partial \ell_{23}}{\partial \theta_{23}}, \frac{\partial \ell}{\partial \theta_{123}} \right)^T. \end{aligned}$$

Note that  $\mathbf{g}(\cdot; \tilde{\boldsymbol{\eta}}) = \mathbf{0}$  holds because  $\tilde{\boldsymbol{\eta}}$  is defined as the solution of  $\mathbf{g}(\cdot; \boldsymbol{\eta}) = \mathbf{0}$ . By using the mean value theorem for  $\mathbf{g}(\cdot; \tilde{\boldsymbol{\eta}})$  at  $\boldsymbol{\eta}_0$ , we have

$$\begin{aligned}\mathbf{0} &= \mathbf{g}(Y_1, \dots, Y_n; \tilde{\boldsymbol{\eta}}) \\ &= \mathbf{g}(Y_1, \dots, Y_n; \boldsymbol{\eta}_0) + J(\tilde{\boldsymbol{\eta}})(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0),\end{aligned}\quad (19)$$

where

$$J(\tilde{\boldsymbol{\eta}}) = \left. \frac{\partial \mathbf{g}(Y_1, \dots, Y_n; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}},$$

and  $\tilde{\boldsymbol{\eta}} \equiv \text{diag}(\alpha_1, \dots, \alpha_7)\tilde{\boldsymbol{\eta}} + \text{diag}(1 - \alpha_1, \dots, 1 - \alpha_7)\boldsymbol{\eta}_0$  for  $(\alpha_1, \dots, \alpha_7) \in [0, 1]^7$ . From Eq. (19), we have  $J(\tilde{\boldsymbol{\eta}})(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = -\mathbf{g}(Y_1, \dots, Y_n; \boldsymbol{\eta}_0)$ . Thus,

$$\sqrt{n}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = -\left[\frac{1}{n}J(\tilde{\boldsymbol{\eta}})\right]^{-1} \frac{1}{\sqrt{n}}\mathbf{g}(\cdot; \boldsymbol{\eta}_0). \quad (20)$$

The first factor on the right-hand side of Eq. (20) can be applied to the law of large numbers because it is divided by the sample size  $n$ . By assuming  $\tilde{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}_0$  for  $n \rightarrow \infty$ ,  $\tilde{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}_0$  also holds for  $n \rightarrow \infty$ . Recalling Eq. (7), we obtain

$$\frac{1}{n}J(\tilde{\boldsymbol{\eta}}) \xrightarrow{p} E\left[\left.\frac{\partial s(\mathbf{Y}; \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0}\right] = D. \quad (21)$$

The second factor can be applied to the multivariate central limit theorem because it is divided by  $\sqrt{n}$ . From Lemma 1 and Eq. (8), we obtain

$$\frac{1}{\sqrt{n}}\mathbf{g}(Y_1, \dots, Y_n; \boldsymbol{\eta}_0) \xrightarrow{D} N(\mathbf{0}, M). \quad (22)$$

Therefore, by applying Slutsky's theorem to Eqs. (21) and (22), the following equation holds.

$$\sqrt{n}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{D} N(\mathbf{0}, D^{-1}M(D^{-1})^T).$$

Hence, the proof is complete.  $\square$

**Acknowledgements** The authors would like to thank the Center for Intelligent Maintenance System, University of Cincinnati, for providing the experimental data. The authors would also like to thank the two anonymous reviewers for their valuable comments that helped to improve the clarity and completeness of this paper.

**Declarations**

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Cossette, H., Cote, M. P., Marceau, E., & Moutanabbir, K. (2013). Multivariate distribution defined with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3), 560–572. <https://doi.org/10.1016/j.insmatheco.2013.03.006>.
- Cramer, H. (1945). *Mathematical methods of statistics*. Princeton: Princeton University Press.
- Eryilmaz, S., & Tank, F. (2012). On reliability analysis of a two-dependent-unit series system with a standby unit. *Applied Mathematics and Computation*, 218(15), 7792–7797. <https://doi.org/10.1016/j.amc.2012.01.046>.
- Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47(3–4), 307–323. <https://doi.org/10.1093/biomet/47.3-4.307>.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292), 698–707. <https://doi.org/10.2307/2281591>.
- Jaworski, P., Durante, F., Härdle, W., & Rychlik, T. (Eds.). (2010). *Copula theory and its applications*. Berlin: Springer.
- Joe, H., & Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. *Technical Report, Department of Statistics, University of British Columbia*, 166, 1–21. <https://doi.org/10.14288/1.0225985>.
- Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2), 401–419. <https://doi.org/10.1016/j.jmva.2004.06.003>.
- Joe, H. (2014). *Dependence Modeling with Copulas*. London: Chapman & Hall.
- Johnson, N. L., & Kotz, S. (1975). On some generalized Farlie–Gumbel–Morgenstern distributions. *Communications in Statistics*, 4(5), 415–427. <https://doi.org/10.1080/03610927508827258>.
- Johnson, N. L., & Kotz, S. (1977). On some generalized Farlie–Gumbel–Morgenstern distributions-II regression, correlation and further generalizations. *Communications in Statistics—Theory and Methods*, 6(6), 485–496. <https://doi.org/10.1080/03610927708827509>.
- Online document Lee, J., Qiu, H., Yu, G., Lin, J. & Rexnord Technical Services. (2007). Bearing Data Set. IMS, University of Cincinnati. NASA Ames Prognostics Data Repository. <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>. Accessed 1 2021.
- Lehmann, E. L., & Casella, G. (1998). *Theory of Point Estimation* (2nd ed.). New York: Springer.
- MacDonald, I. L. (2014). Does Newton–Raphson really fail? *Statistical Methods in Medical Research*, 23(3), 308–311. <https://doi.org/10.1177/0962280213497329>.
- McCool, J. I. (2012). Testing for dependency of failure times in life testing. *Technometrics*, 48(1), 41–48. <https://doi.org/10.1198/004017005000000544>.
- Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsblatt für Mathematische Statistik*, 8, 234–235.
- Navarro, J., Ruiz, J. M., & Sandoval, C. J. (2007). Properties of coherent systems with dependent components. *Communications in Statistics—Theory and Methods*, 36, 175–191. <https://doi.org/10.1080/03610920600966316>.
- Navarro, J., & Durante, F. (2017). Copula-based representations for the reliability of the residual lifetimes of coherent systems with dependent components. *Journal of Multivariate Analysis*, 158, 87–102. <https://doi.org/10.1016/j.jmva.2017.04.003>.
- Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). New York: Springer.
- Patton, A. J. (2006). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21(2), 147–173. <https://doi.org/10.1002/jae.865>.
- Qiu, H., Lee, J., Lin, J., & Yuc, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, 289, 1066–1090. <https://doi.org/10.1016/j.jsv.2005.03.007>.

- Shih, J. H., Chang, Y. T., Konno, Y., & Emura, T. (2019). Estimation of a common mean vector in bivariate meta-analysis under the FGM copula. *Statistics*, *53*(3), 673–695. <https://doi.org/10.1080/02331888.2019.1581782>.
- Xu, J. J. (1996). Statistical modeling and inference for multivariate and longitudinal discrete response data. Ph.D. thesis, Department of Statistics, University of British Columbia.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.