STATISTICS
2019, VOL. 53, NO. 3, 673–695
https://doi.org/10.1080/02331888.2019.1581782

Estimation of a common mean vector in bivariate meta-analysis under the FGM copula

Jia-Han Shih[a], Yoshihiko Konno[b], Yuan-Tsung Chang[c] and Takeshi Emura[a]

[a]Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan; [b]Department of Mathematical and Physical Sciences, Japan Women's University, Tokyo, Japan; [c]Department of Social Information, Faculty of Studies on Contemporary Society, Mejiro University, Tokyo, Japan

ABSTRACT
We propose a bivariate Farlie–Gumbel–Morgenstern (FGM) copula model for bivariate meta-analysis, and develop a maximum likelihood estimator for the common mean vector. With the aid of novel mathematical identities for the FGM copula, we derive the expression of the Fisher information matrix. We also derive an approximation formula for the Fisher information matrix, which is accurate and easy to compute. Based on the theory of independent but not identically distributed (i.n.i.d.) samples, we examine the asymptotic properties of the estimator. Simulation studies are given to demonstrate the performance of the proposed method, and a real data analysis is provided to illustrate the method.

ARTICLE HISTORY
Received 30 January 2018
Accepted 20 January 2019

KEYWORDS
Asymptotic theory; copula; Fisher information; maximum likelihood estimation; multivariate analysis; Stein's identity

1. Introduction

Multivariate meta-analysis has been widely applied to scientific areas such as education and medicine, where multiple outcomes are measured across different studies. A well-known example in educational research is a meta-analysis of bivariate test scores (on verbal and mathematics) collected from different studies [1,2]. Section 7 shall introduce our original data on bivariate entrance examination scores (on mathematics and statistics) obtained across 5 different academic years (from 2013 to 2017).

If one performs separate univariate meta-analyses on multivariate outcomes, any possible dependence is ignored. Riley [2] has shown that ignoring dependence between outcomes increases the mean-square error for estimating parameters. In the example of the bivariate test scores, positive dependence arises due to students' intellectual ability. In medical research, positive dependence between two survival outcomes is a key to validate surrogacy [3,4] and to predict overall survival [5,6].

One should consider multivariate meta-analysis to perform simultaneous analyses without ignoring dependence between outcomes. Multivariate analyses would increase the

CONTACT Takeshi Emura takeshiemura@gmail.com Graduate Institute of Statistics, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan City 32001, Taiwan

Supplemental data for this article can be accessed here. https://doi.org/10.1080/02331888.2019.1581782

efficiency of parameter estimation and allow one to study dependence patterns among outcomes, provided that the form of the dependence pattern is specified correctly. Unfortunately, the developments of multivariate meta-analysis are limited to the multivariate normal model [2,7,8]. This motivates us to explore an alternative model giving a different dependence pattern from the multivariate (bivariate) normal distribution model.

The Farlie–Gumbel–Morgenstern (FGM) model provides the classical way to consider a dependence pattern between two random variables, in a different manner from the bivariate normal model. The FGM model was first introduced by Morgenstern [9], which can even be traced back to Eyraud [10]. It was later studied by Farlie [11] and Gumbel [12]. The FGM distribution is naturally derived as a linear combination of products of the distribution of order statistics [13]. More general constructions based on order statistics were studied by various authors [14–16].

The FGM copula is a copula derived from the FGM models. Due to its simple form, the FGM models have been used to demonstrate both theoretical and practical aspects of copulas. Genest and Favre [17] adopted the FGM copula as an example to demonstrate modelling strategies, rank-based inference procedures, and goodness-of-fit tests on copulas. These demonstrations may not be possible under the normal copula, a copula derived from the bivariate normal distribution. In applications to the medical study, Kim et al. [18] analysed the directional dependence of genes by using the FGM type copulas. Martinez and Achcar [19] applied the FGM copula with the cure fraction model to analyse two real data from the cervical cancer study and the diabetic retinopathy study by Bayesian approaches.

This paper aims to consider the problem of estimating a common mean by fitting the FGM copula model for dependence between two normally distributed variates. While the FGM copula has been extensively studied in the literature, its real applications are relatively scarce, compared to more common copulas such as the Clayton copula and the normal copula. Throughout this paper, however, we shall stress that the FGM copula has several mathematically important features that make it suitable for the problem of estimating a common mean in a fixed-effects meta-analysis.

In this paper, we adopt maximum likelihood estimation for the common mean vector. We derive new mathematical identities unique to the FGM model to obtain the expression of the Fisher information matrix. We also derive a linear approximation to the Fisher information matrix, which is accurate and easy to compute. We examine the asymptotic properties of the estimator based on the theory of independent but not identically distributed (i.n.i.d.) samples.

This paper is organized as follows. Section 2 reviews the common mean bivariate normal model. Section 3 proposes the model, identities, and estimator. Section 4 derives the Fisher information matrix. Section 5 develops the asymptotic theory. Section 6 conducts simulations. Section 7 performs data analysis. Section 8 summarizes the paper and gives discussions for future works.

2. Common mean bivariate normal model

We introduce a common mean model for bivariate meta-analysis which is discussed by various authors (e.g., [7,8]). For each study i, consider a random vector $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^{\mathrm{T}}$

following a bivariate normal distribution.

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} \sim \text{BVN}\left(\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \mathbf{C}_i = \begin{bmatrix} \sigma_{i1}^2 & \rho_i\sigma_{i1}\sigma_{i2} \\ \rho_i\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{bmatrix}\right), \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\rho_i \in (-1, 1)$ is the within-study correlation for study i. We call $\boldsymbol{\mu}$ 'common mean vector' since it is common across $i = 1, 2, \ldots, n$ ([20]; Example 2.2.1, [21]).

Meta-analysis is a method to combine the known results from several independent studies and to make inference for a population. The common mean model is suitable for a fixed-effects meta-analysis, where the ith study contains estimate $\mathbf{Y}_i$ of the target parameter $\boldsymbol{\mu}$ with its covariance matrix $\mathbf{C}_i$. In meta-analyses, $\mathbf{C}_i$'s are assumed known. Hence, $\mathbf{Y}_i$'s are independent but not identically distributed (i.n.i.d.) across $i = 1, 2, \cdots, n$.

The problem of estimating the common mean in model (1) has been separately studied between a decision-theoretic framework and a meta-analytic framework. In the decision theory, $\mathbf{C}_i$'s are assumed unknown [20]. In meta-analyses, $\mathbf{C}_i$'s are often known from published data or summary data [8]. In both frameworks, the bivariate normal distribution is exclusively applied for analysis.

Note that the common mean bivariate normal model (1) is represented as

$$\begin{aligned} \Pr(Y_{i1} \le y_1, Y_{i2} \le y_2) &= \Phi_{\rho_i}\left(\frac{y_1 - \mu_1}{\sigma_{i1}}, \frac{y_2 - \mu_2}{\sigma_{i2}}\right) \\ &= C_{\rho_i}^{\text{Normal}}\left(\Phi\left(\frac{y_1 - \mu_1}{\sigma_{i1}}\right), \Phi\left(\frac{y_2 - \mu_2}{\sigma_{i2}}\right)\right), \end{aligned}$$

where $\Phi_\rho(\cdot, \cdot)$ is the bivariate cumulative distribution function (c.d.f.) of the bivariate standard normal distribution with correlation ρ, $\Phi(\cdot)$ is the c.d.f. of $N(0, 1)$, and

$$C_\rho^{\text{Normal}}(u, v) = \Phi_\rho\{\Phi^{-1}(u), \Phi^{-1}(v)\}, \quad 0 \le u, v \le 1.$$

The function $C_\rho^{\text{Normal}} : [0, 1]^2 \mapsto [0, 1]$ is called 'normal copula'. By replacing C_ρ^{Normal} with any other copula, one can create a non-bivariate normal distribution [17,22].

3. Proposed methods

This section introduces the proposed model, some mathematical identities, and estimation method.

3.1. The bivariate FGM model

As an alternative to the normal copula, we consider the FGM copula defined as

$$C_\theta^{\text{FGM}}(u, v) = uv\{1 + \theta(1 - u)(1 - v)\}, \quad 0 \le u, v \le 1,$$

where $\theta \in [-1, 1]$ is the dependence parameter. The form of C_θ^{FGM} is simpler than the form of C_ρ^{Normal}. Under the FGM copula with the uniform margins on the unit interval $[0, 1]$, the relationship between θ and the correlation ρ is $\theta = 3\rho$ [23]. This relationship implies that the correlation is restricted to $\rho \in [-1/3, 1/3]$ for the FGM copula.

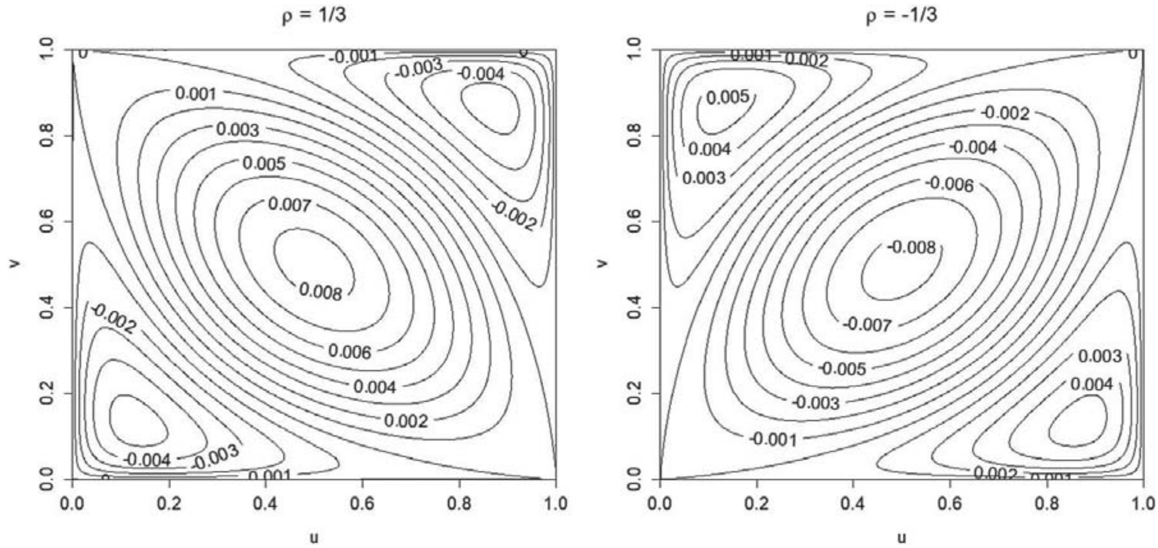

Figure 1. The difference $C^{\mathrm{FGM}}_{3\rho} - C^{\mathrm{Normal}}_{\rho}$ with correlation $\rho = 1/3$ or $-1/3$.

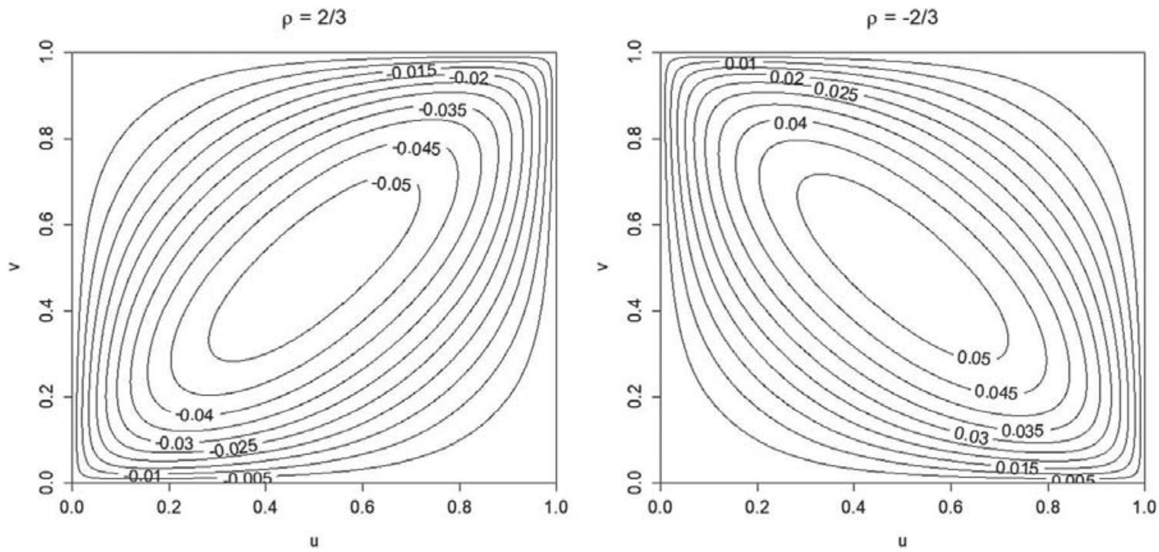

Figure 2. The difference $C^{\mathrm{FGM}}_{\min[1,\max\{-1,3\rho\}]} - C^{\mathrm{Normal}}_{\rho}$ with correlation $\rho = 2/3$ or $-2/3$.

Figure 1 shows that the difference $C^{\mathrm{FGM}}_{3\rho} - C^{\mathrm{Normal}}_{\rho}$ is nonzero even when the two copulas has the same correlation $\rho = 1/3$ or $\rho = -1/3$. Under $\rho = 1/3$, we observe that $C^{\mathrm{FGM}}_{3\rho}(u,u) > C^{\mathrm{Normal}}_{\rho}(u,u)$ for $u \approx 0.5$ and that $C^{\mathrm{FGM}}_{3\rho}(u,u) < C^{\mathrm{Normal}}_{\rho}(u,u)$ for $u \approx 0.1$ and $u \approx 0.9$. A similar pattern can be found under $\rho = -1/3$.

The FGM copula cannot yield a strong positive correlation $(1/3 < \rho < 1)$ or strong negative correlation $(-1 < \rho < -1/3)$. To compare the FGM and normal copulas, we reparametrize the FGM copula as $C^{\mathrm{FGM}}_{\min[1,\max\{-1,3\rho\}]}$ so that $-1 \le \min[1, \max\{-1, 3\rho\}] \le 1$ for $-1 < \rho < 1$. This is a boundary correction similar to Genest and Nešlehová [22]. Figure 2 again shows that the difference $C^{\mathrm{FGM}}_{\min[1,\max\{-1,3\rho\}]} - C^{\mathrm{Normal}}_{\rho}$ is nonzero when $\rho = 2/3$ or $\rho = -2/3$. We observe that $C^{\mathrm{FGM}}_{\min[1,\max\{-1,3\rho\}]} < C^{\mathrm{Normal}}_{\rho}$ under $\rho = 2/3$ and that $C^{\mathrm{FGM}}_{\min[1,\max\{-1,3\rho\}]} > C^{\mathrm{Normal}}_{\rho}$ under $\rho = -2/3$.

Applying the FGM copula to the normal margins, we define the bivariate FGM model

$$\Pr(Y_{i1} \le y_1, Y_{i2} \le y_2) = \Phi\left(\frac{y_1-\mu_1}{\sigma_{i1}}\right)\Phi\left(\frac{y_2-\mu_2}{\sigma_{i2}}\right) \times \left[1+\theta_i\left\{1-\Phi\left(\frac{y_1-\mu_1}{\sigma_{i1}}\right)\right\}\left\{1-\Phi\left(\frac{y_2-\mu_2}{\sigma_{i2}}\right)\right\}\right], \quad (2)$$

where the marginal distributions are $Y_{i1} \sim N(\mu_1, \sigma_{i1}^2)$ and $Y_{i2} \sim N(\mu_2, \sigma_{i2}^2)$. Equation (2) is a valid distribution function [22] whose joint density is

$$f_{i,\mu}(\mathbf{y}) = \frac{1}{2\pi\sigma_{i1}\sigma_{i2}}\exp\left\{-\frac{(y_1-\mu_1)^2}{2\sigma_{i1}^2}-\frac{(y_2-\mu_2)^2}{2\sigma_{i2}^2}\right\}Q_i(\mathbf{y};\boldsymbol{\mu}), \quad \mathbf{y}\in\mathbb{R}^2, \quad (3)$$

where $\mathbf{y} = (y_1, y_2)^{\mathrm{T}}$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^{\mathrm{T}}$, and

$$Q_i(\mathbf{y};\boldsymbol{\mu}) = 1+\theta_i\left\{1-2\Phi\left(\frac{y_1-\mu_1}{\sigma_{i1}}\right)\right\}\left\{1-2\Phi\left(\frac{y_2-\mu_2}{\sigma_{i2}}\right)\right\}, \quad \theta_i \in [-1, 1].$$

3.2. Identities

We have the following identities, which are useful for subsequent discussions.

Lemma 3.1: *For integrable functions g_1 and g_2,*

$$E_{\boldsymbol{\mu}}\left\{\frac{g_1(Y_{i1})g_2(Y_{i2})}{Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}\right\} = E_{\mu_1}\{g_1(Y_{i1})\}E_{\mu_2}\{g_2(Y_{i2})\},$$

$$E_{\boldsymbol{\mu}}\{g_1(Y_{i1})g_2(Y_{i2})\} = E_{\mu_1}\{g_1(Y_{i1})\}E_{\mu_2}\{g_2(Y_{i2})\} + \theta_i E_{\mu_1}\left[g_1(Y_{i1})\left\{1-2\Phi\left(\frac{Y_{i1}-\mu_1}{\sigma_{i1}}\right)\right\}\right] \times E_{\mu_2}\left[g_2(Y_{i2})\left\{1-2\Phi\left(\frac{Y_{i2}-\mu_2}{\sigma_{i2}}\right)\right\}\right].$$

Lemma 3.1 is proven by calculating the left-hand sides using the formula of $f_{i,\boldsymbol{\mu}}$.

A striking feature of Lemma 3.1 is that a cross-moment for Y_{i1} and Y_{i2} reduces to the marginal moments of $Y_{i1} \sim N(\mu_1, \sigma_{i1}^2)$ and $Y_{i2} \sim N(\mu_2, \sigma_{i2}^2)$, respectively. This is a unique mathematical property of the FGM copula, which is not applicable to other copulas.

Lemma 3.1 leads to the following identities related to the derivatives of $Q_i(\mathbf{y};\boldsymbol{\mu})$.

Lemma 3.2: *For $i\ (= 1, 2, \cdots, n)$ and $j = 1, 2$,*

$$E_{\boldsymbol{\mu}}\left\{\left(\frac{Y_{i1}-\mu_1}{\sigma_{i1}^2}\right)\left(\frac{Y_{i2}-\mu_2}{\sigma_{i2}^2}\right)\right\} = E_{\boldsymbol{\mu}}\left\{\frac{1}{Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}\frac{\partial^2 Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}{\partial\mu_1\partial\mu_2}\right\} = \frac{\theta_i}{\pi\sigma_{i1}\sigma_{i2}},$$

$$E_{\mu}\left[\frac{Y_{ij}-\mu_j}{\sigma_{ij}^2}\left\{\frac{1}{Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}\frac{\partial Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}{\partial \mu_{3-j}}\right\}\right]=-\frac{\theta_i}{\pi\sigma_{i1}\sigma_{i2}},$$

$$E_{\mu}\left\{\frac{1}{Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}\frac{\partial^2 Q_i(\mathbf{Y}_i;\boldsymbol{\mu})}{\partial \mu_j^2}\right\}=0,$$

$$\frac{\partial^2 Q_i(\boldsymbol{y};\boldsymbol{\mu})}{\partial\mu_1\partial\mu_2}=\frac{4\theta_i}{\sigma_{i1}\sigma_{i2}}\phi\left(\frac{y_1-\mu_1}{\sigma_{i1}}\right)\phi\left(\frac{y_2-\mu_2}{\sigma_{i2}}\right),$$

$$\frac{\partial Q_i(\boldsymbol{y};\boldsymbol{\mu})}{\partial\mu_j}=\frac{2\theta_i}{\sigma_{ij}}\phi\left(\frac{y_j-\mu_j}{\sigma_{ij}}\right)\left\{1-2\Phi\left(\frac{y_{3-j}-\mu_{3-j}}{\sigma_{i(3-j)}}\right)\right\},$$

$$\frac{\partial^2 Q_i(\boldsymbol{y};\boldsymbol{\mu})}{\partial\mu_j^2}=\frac{2\theta_i(y_j-\mu_j)}{\sigma_{ij}^3}\phi\left(\frac{y_j-\mu_j}{\sigma_{ij}}\right)\times\left\{1-2\Phi\left(\frac{y_{3-j}-\mu_{3-j}}{\sigma_{i(3-j)}}\right)\right\},$$

where $\phi(\cdot)$ denotes the density of $N(0,1)$.

These identities shall be useful for deriving the Fisher information matrix. All of them are derived by first applying Lemma 3.1 with some g_1 and g_2 and then calculating the marginal moments of Y_{i1} and Y_{i2}. The detailed derivation is available in Supplementary Material.

The FGM model has been widely discussed in the literature, partly due to its elegant mathematical properties. It provides closed-form expressions for various dependence measures. For example, Kendall's tau and Spearman's rho are

$$\tau_K=4\int_0^1\int_0^1 C_\theta^{\mathrm{FGM}}(u,v)dC_\theta^{\mathrm{FGM}}(u,v)-1=\frac{2\theta}{9},$$

$$\rho_S=12\int_0^1\int_0^1 C_\theta^{\mathrm{FGM}}(u,v)dudv-3=\frac{\theta}{3}.$$

They are measures of dependence free from the marginal distributions (Examples 5.2 and 5.7, [24]). In addition, the Pearson correlation has closed-form expressions under several well-known marginal distributions. For instance, the Pearson correlation under the normal and exponential margins are θ/π and $\theta/4$, respectively [23]. One may directly obtain the case of normal margins by applying the first identity in Lemma 3.2.

3.3. Maximum likelihood estimation

This section proposes a maximum likelihood estimator for the common mean vector under the FGM model in Equation (2). Suppose $\mathbf{Y}_i=(Y_{i1},Y_{i2})^{\mathrm{T}}, i=1,2,\cdots,n$ are independent samples with the joint density defined in Equation (3). Then, the log-likelihood is

$$\ell_n(\boldsymbol{\mu}) = \log L_n(\boldsymbol{\mu}) = \text{constant} - \sum_{i=1}^{n} \frac{(Y_{i1} - \mu_1)^2}{2\sigma_{i1}^2} - \sum_{i=1}^{n} \frac{(Y_{i2} - \mu_2)^2}{2\sigma_{i2}^2} + \sum_{i=1}^{n} \log Q_i(\mathbf{Y}_i; \boldsymbol{\mu}).$$

The MLE of $\boldsymbol{\mu}$ is defined as $\hat{\boldsymbol{\mu}}_n^{\text{MLE}} = \arg\max_{\boldsymbol{\mu}\in\Theta} \ell_n(\boldsymbol{\mu})$, where $\Theta \subset \mathbb{R}^2$ is a parameter space. It is obtained by solving $\partial \ell_n(\boldsymbol{\mu})/\partial \mu_j = 0$ for $j = 1, 2$ which are equivalent to

$$\mu_j = \left(\sum_{i=1}^{n} \frac{1}{\sigma_{ij}^2} \right)^{-1} \sum_{i=1}^{n} \left(\frac{Y_{ij}}{\sigma_{ij}^2} + \frac{1}{Q_i(\mathbf{Y}_i; \boldsymbol{\mu})} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_j} \right), \quad j = 1, 2.$$

It is clear that $\theta_i = 0$ implies $Q_i(\mathbf{y}; \boldsymbol{\mu}) = 1$ and $\partial Q_i(\mathbf{y}; \boldsymbol{\mu})/\partial \mu_1 = \partial Q_i(\mathbf{y}; \boldsymbol{\mu})/\partial \mu_2 = 0$. Therefore, if $\theta_i = 0$ for $i = 1, 2, \cdots, n$, one can obtain the univariate estimators

$$\mu_j^{(0)} = \left(\sum_{i=1}^{n} \frac{1}{\sigma_{ij}^2} \right)^{-1} \sum_{i=1}^{n} \frac{Y_{ij}}{\sigma_{ij}^2}, \quad j = 1, 2.$$

This is the usual estimator of a single common mean (Example 2.2.2, [21]). Since $Q_i(\mathbf{y}; \boldsymbol{\mu})$ depends nonlinearly on $\boldsymbol{\mu}$ for $\theta_i \neq 0$, the closed-form expression for $\hat{\boldsymbol{\mu}}_n^{\text{MLE}}$ is not available. Thus, we suggest applying the Newton–Raphson (NR) algorithm to obtain the MLE. A concrete algorithm (Algorithm S1) is given in Supplementary Material.

4. Fisher information matrix

This section studies the Fisher information matrix. It contains all necessary information about the asymptotic distributions of $\hat{\boldsymbol{\mu}}_n^{\text{MLE}}$, and is useful to compute the standard error and confidence region. The derivatives of the log-density are

$$\frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{y})}{\partial \mu_j} = \frac{y_j - \mu_j}{\sigma_{ij}^2} + \frac{1}{Q_i(\mathbf{y}; \boldsymbol{\mu})} \frac{\partial Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_j}, \quad j = 1, 2,$$

$$\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{y})}{\partial \mu_j^2} = \frac{1}{Q_i(\mathbf{y}; \boldsymbol{\mu})} \frac{\partial^2 Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_j^2} - \left\{ \frac{1}{Q_i(\mathbf{y}; \boldsymbol{\mu})} \frac{\partial Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_j} \right\}^2 - \frac{1}{\sigma_{ij}^2}, \quad j = 1, 2,$$

$$\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{y})}{\partial \mu_1 \partial \mu_2} = \frac{1}{Q_i(\mathbf{y}; \boldsymbol{\mu})} \frac{\partial^2 Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_1 \partial \mu_2} - \frac{1}{Q_i(\mathbf{y}; \boldsymbol{\mu})^2} \frac{\partial Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_1} \frac{\partial Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_2}.$$

We define the 2×2 Fisher information matrix $I_i(\boldsymbol{\mu})$ for $i\ (= 1, 2, \cdots, n)$ as

$$I_{i,jk}(\boldsymbol{\mu}) = E_{\boldsymbol{\mu}} \left\{ \frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_j} \frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_k} \right\}, \quad j, k = 1, 2.$$

Lemma 4.1: *For each $i\ (= 1, 2, \cdots, n)$,*

$$E_{\boldsymbol{\mu}} \left\{ \frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_j} \frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_k} \right\} = E_{\boldsymbol{\mu}} \left\{ -\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_j \partial \mu_k} \right\}, \quad j, k = 1, 2.$$

If $f_{i,\mu}(\cdot)$ were a member of the exponential family, Lemma 4.1 would be trivial. In our setting, $f_{i,\mu}(\cdot)$ does not belong to the exponential family unless $\theta_i = 0$. The proof of Lemma 4.1 relies on Lemma 3.2, and is available in Supplementary Material. Lemma 4.1 implies

$$I_i(\boldsymbol{\mu}) = E_{\boldsymbol{\mu}} \begin{Bmatrix} -\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_1^2} & -\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_1 \partial \mu_2} \\ -\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_1 \partial \mu_2} & -\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_2^2} \end{Bmatrix}. \tag{4}$$

The following theorem derives the form of the Fisher information matrix $I_i(\boldsymbol{\mu})$.

Theorem 4.1: *The Fisher information matrix for i $(= 1, 2, \cdots, n)$ does not depend on $\boldsymbol{\mu}$. In addition, it can be decomposed into the sum of the Fisher information matrix for the independent model and the additional information for the dependence parameter θ_i,*

$$I_i = \begin{bmatrix} \frac{1}{\sigma_{i1}^2} & 0 \\ 0 & \frac{1}{\sigma_{i2}^2} \end{bmatrix} + \begin{bmatrix} \frac{4\theta_i^2}{\sigma_{i1}^2} E_i & \frac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}} F_i - \frac{\theta_i}{\pi \sigma_{i1}\sigma_{i2}} \\ \frac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}} F_i - \frac{\theta_i}{\pi \sigma_{i1}\sigma_{i2}} & \frac{4\theta_i^2}{\sigma_{i2}^2} E_i \end{bmatrix},$$

where $\theta_i \in [-1, 1]$,

$$E_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\phi^3(u)\{1 - 2\Phi(v)\}^2 \phi(v)}{1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}} dudv,$$

$$F_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\phi^2(u)\{1 - 2\Phi(u)\}\phi^2(v)\{1 - 2\Phi(v)\}}{1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}} dudv.$$

Proof of Theorem 4.1: By Lemma 4.1, the element $I_{i,12}(\boldsymbol{\mu})$ can be computed as

$$\begin{aligned} E_{\boldsymbol{\mu}} \left\{ -\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial \mu_1 \partial \mu_2} \right\} &= -E_{\boldsymbol{\mu}} \left\{ \frac{1}{Q_i(\mathbf{Y}_i; \boldsymbol{\mu})} \frac{\partial^2 Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_1 \partial \mu_2} \right\} \\ &\quad + E_{\boldsymbol{\mu}} \left\{ \frac{1}{Q_i(\mathbf{Y}_i; \boldsymbol{\mu})^2} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_1} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_2} \right\} \\ &= -\frac{\theta_i}{\pi \sigma_{i1}\sigma_{i2}} + E_{\boldsymbol{\mu}} \left\{ \frac{1}{Q_i(\mathbf{Y}_i; \boldsymbol{\mu})^2} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_1} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_2} \right\}, \end{aligned}$$

where the last equality follows from Lemma 3.2. By straightforward calculations,

$$\begin{aligned} &E_{\boldsymbol{\mu}} \left\{ \frac{1}{Q_i(\mathbf{Y}_i; \boldsymbol{\mu})^2} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_1} \frac{\partial Q_i(\mathbf{Y}_i; \boldsymbol{\mu})}{\partial \mu_2} \right\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{Q_i(\mathbf{y}; \boldsymbol{\mu})^2} \frac{\partial Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_1} \frac{\partial Q_i(\mathbf{y}; \boldsymbol{\mu})}{\partial \mu_2} \frac{1}{\sigma_{i1}\sigma_{i2}} \end{aligned}$$

$$\times \phi\left(\frac{y_1-\mu_1}{\sigma_{i1}}\right)\phi\left(\frac{y_2-\mu_2}{\sigma_{i2}}\right)Q_i(\mathbf{y};\boldsymbol{\mu})dy_1dy_2$$
$$=\frac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\phi^2(u)\{1-2\Phi(u)\}\phi^2(v)\{1-2\Phi(v)\}}{1+\theta_i\{1-2\Phi(u)\}\{1-2\Phi(v)\}}dudv.$$

Thus, we obtain

$$E_{\boldsymbol{\mu}}\left\{-\frac{\partial^2 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial\mu_1\partial\mu_2}\right\}$$
$$=\frac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\phi^2(u)\{1-2\Phi(u)\}\phi^2(v)\{1-2\Phi(v)\}}{1+\theta_i\{1-2\Phi(u)\}\{1-2\Phi(v)\}}dudv-\frac{\theta_i}{\pi\sigma_{i1}\sigma_{i2}}$$
$$=\frac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}}F_i-\frac{\theta_i}{\pi\sigma_{i1}\sigma_{i2}}.$$

The elements $I_{i,11}(\boldsymbol{\mu})$ and $I_{i,22}(\boldsymbol{\mu})$ are derived in a similar fashion. After obtaining all elements of $I_i(\boldsymbol{\mu})$, the decomposition is obvious. ∎

Theorem 4.1 helps us to interpret the role of the dependence parameter θ_i on the Fisher information matrix. The decomposition in Theorem 4.1 is a natural result since the FGM copula can be written as the sum of the independent copula and a function of θ_i, namely

$$C_{\theta_i}^{\mathrm{FGM}}(u,v)=uv+\theta_i u(1-u)v(1-v).$$

To show that the double integrals in Theorem 4.1 are finite, we need the following lemma.

Lemma 4.2: *The following inequalities hold* $\frac{u}{u^2+1}\phi(u)<1-\Phi(u)<\frac{1}{u}\phi(u)$ *if* $u>0$; $\frac{|u|}{u^2+1}\phi(u)<\Phi(u)<\frac{1}{|u|}\phi(u)$ *if* $u<0$.

The inequalities in Lemma 4.2 follow from Example 3.6.3 and Exercise 3.47 in Casella and Berger [25].

Theorem 4.2: *For all* $\theta_i\in[-1,1]$, E_i *and* F_i *are finite.*

Proof of Theorem 4.2: We first show that E_i is finite. Since the function inside the double integral is always positive, it suffices to show that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\phi^3(u)\{1-2\Phi(v)\}^2\phi(v)}{1+\theta_i\{1-2\Phi(u)\}\{1-2\Phi(v)\}}dudv<\infty,\quad \text{for } \theta_i\in[-1,1]. \tag{5}$$

The proof is separated into three different cases: $\theta_i=1,-1$, and $\theta_i\in(-1,1)$.

If $\theta_i \in (-1, 1)$, we have $1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\} \geq 1 - |\theta_i| > 0$. Then,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\phi^3(u)\{1 - 2\Phi(v)\}^2\phi(v)}{1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}} dudv$$

$$\leq \frac{1}{1 - |\theta_i|} \int_{-\infty}^{\infty} \phi^3(u)du \int_{-\infty}^{\infty} \{1 - 2\Phi(v)\}^2\phi(v)dv$$

$$= \frac{1}{6\sqrt{3}\pi(1 - |\theta_i|)} < \infty.$$

Hence we have shown the double integral in Equation (5) is finite if $\theta_i \in (-1, 1)$. The proof for the case of $\theta_i = 1 \text{or} - 1$ requires Lemma 4.2 and is given in Supplementary Material. The remaining proof for F_i will be obtained automatically in the proof of the next theorem. ■

Theorem 4.3: *For all $\theta_i \in [-1, 1]$, the determinant of I_i can be expressed as*

$$\det(I_i) = \frac{16\theta_i^4}{\sigma_{i1}^2\sigma_{i2}^2}(E_i^2 - F_i^2) + \frac{8\theta_i^2}{\sigma_{i1}^2\sigma_{i2}^2}\left(E_i + \frac{\theta_i}{\pi}F_i\right) + \frac{1}{\sigma_{i1}^2\sigma_{i2}^2}\left(1 - \frac{\theta_i^2}{\pi^2}\right).$$

In addition, $\det(I_i)$ *is positive and I_i is positive definite.*

Proof of Theorem 4.3: The expression of $\det(I_i)$ follows from straightforward calculations. If $E_i \geq |F_i|$, then $E_i^2 \geq F_i^2$ and $E_i > |\theta_i F_i|/\pi$. Therefore, $E_i \geq |F_i|$ implies $\det(I_i) > 0$ since $1 - \theta_i^2/\pi^2 > 0$. Thus, it suffices to show the inequality $E_i \geq |F_i|$. It holds according to the Cauchy–Schwarz inequality as follow:

$$|F_i| \leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\phi^2(u)|1 - 2\Phi(u)|\phi^2(v)|1 - 2\Phi(v)|}{1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}} dudv$$

$$\leq \left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\phi^{3/2}(u)|1 - 2\Phi(v)|\phi^{1/2}(v)}{[1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}]^{1/2}}\right)^2 dudv\right\}^{1/2}$$

$$\times \left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\phi^{3/2}(v)|1 - 2\Phi(u)|\phi^{1/2}(u)}{[1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}]^{1/2}}\right)^2 dudv\right\}^{1/2}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\phi^3(u)\{1 - 2\Phi(v)\}^2\phi(v)}{1 + \theta_i\{1 - 2\Phi(u)\}\{1 - 2\Phi(v)\}} dudv = E_i.$$

The above inequality not only ensures $\det(I_i) > 0$ but also guarantees $|F_i| < \infty$. Since $4\theta_i^2 E_i/\sigma_{i1}^2 + 1/\sigma_{i1}^2 > 0$, we have shown that the upper left 1×1 and 2×2 determinants of I_i are positive. Hence I_i is positive definite. ■

The expressions for E_i and F_i in Theorem 4.1 are still difficult to evaluate in practice. In order to make these expressions more tractable, we apply the Taylor expansion for $[1+\theta_i\{1-2\Phi(u)\}\{1-2\Phi(v)\}]^{-1}$. Without loss of generality, we exclude the cases of $\theta_i = 1$ and -1 by assuming $\theta_i \in (-1, 1)$. Based on this subtle assumption, we have $-1 < (-\theta_i)^k\{1-2\Phi(u)\}^k\{1-2\Phi(v)\}^k < 1$, for all u, v. Then,

$$\frac{1}{1+\theta_i\{1-2\Phi(u)\}\{1-2\Phi(v)\}} = \sum_{k=0}^{\infty}(-\theta_i)^k\{1-2\Phi(u)\}^k\{1-2\Phi(v)\}^k.$$

After some calculations, we obtain the new expressions of E_i and F_i.

Theorem 4.4: *If $\theta_i \in (-1, 1)$, then E_i and F_i in Theorem 4.1 have alternative forms*

$$E_i = \sum_{k=0}^{\infty}\frac{\theta_i^{2k}}{2k+3}\sum_{\ell=0}^{2k}\binom{2k}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^3(u)\Phi^\ell(u)du,$$

$$F_i = -\sum_{k=0}^{\infty}\theta_i^{2k+1}\left\{\sum_{\ell=0}^{2k+2}\binom{2k+2}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^2(u)\Phi^\ell(u)du\right\}^2.$$

The proof of Theorem 4.4 involves the Lebesgue dominated convergence theorem that ensures the interchangeability between integration and infinite summation, and the binomial theorem. The proof of Theorem 4.4 is given in Supplementary Material.

Since $|\theta_i| < 1$, the contributions of high-order terms in the expressions of E_i and F_i are small for large k. It turns out that only $k = 0$ and 1 are essential and $k \geq 2$ are negligible. Accordingly, we let $E_i = A_1 + A_2\theta_i^2 + e_i \equiv \tilde{E}_i + e_i$, where $A_1 \equiv \sqrt{3}/18\pi$,

$$A_2 \equiv \frac{1}{5}\sum_{\ell=0}^{2}\binom{2}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^3(u)\Phi^\ell(u)du,$$

$$e_i \equiv \sum_{k=2}^{\infty}\frac{\theta_i^{2k}}{2k+3}\sum_{\ell=0}^{2k}\binom{2k}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^3(u)\Phi^\ell(u)du.$$

Note that e_i is the remainder terms of E_i. Similarly, we let $F_i = B_1\theta_i + B_2\theta_i^3 + f_i \equiv \tilde{F}_i + f_i$, where

$$B_1 \equiv -\left\{\sum_{\ell=0}^{2}\binom{2}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^2(u)\Phi^\ell(u)du\right\}^2,$$

$$B_2 \equiv -\left\{\sum_{\ell=0}^{4}\binom{4}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^2(u)\Phi^\ell(u)du\right\}^2,$$

$$f_i \equiv -\sum_{k=2}^{\infty}\theta_i^{2k+1}\left\{\sum_{\ell=0}^{2k+2}\binom{2k+2}{\ell}(-2)^\ell\int_{-\infty}^{\infty}\phi^2(u)\Phi^\ell(u)du\right\}^2.$$

One can obtain the values of A_1, A_2, B_1, and B_2 numerically. Finally, we approximate the exact Fisher information by its linear approximation.

Theorem 4.5: (Approximate Fisher information matrix) *If* $\theta_i \in (-1, 1)$, *then*

$$I_i = \begin{bmatrix} \dfrac{1}{\sigma_{i1}^2} & 0 \\ 0 & \dfrac{1}{\sigma_{i2}^2} \end{bmatrix} + \begin{bmatrix} \dfrac{4\theta_i^2}{\sigma_{i1}^2}\tilde{E}_i & \dfrac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}}\tilde{F}_i - \dfrac{\theta_i}{\pi\sigma_{i1}\sigma_{i2}} \\ \dfrac{4\theta_i^2}{\sigma_{i1}\sigma_{i2}}\tilde{F}_i - \dfrac{\theta_i}{\pi\sigma_{i1}\sigma_{i2}} & \dfrac{4\theta_i^2}{\sigma_{i2}^2}\tilde{E}_i \end{bmatrix} + \begin{bmatrix} e_{i1}^* & f_i^* \\ f_i^* & e_{i2}^* \end{bmatrix},$$

where $\tilde{E}_i = 0.0306 + 0.0030 \times \theta_i^2$, $\tilde{F}_i = -0.0037 \times \theta_i - 0.0008 \times \theta_i^3$, $e_{i1}^* \equiv 4\theta_i^2 e_i/\sigma_{i1}^2 = O(\theta_i^6)$, $e_{i2}^* \equiv 4\theta_i^2 e_i/\sigma_{i2}^2 = O(\theta_i^6)$, *and* $f_i^* \equiv 4\theta_i^2 f_i/(\sigma_{i1}\sigma_{i2}) = O(\theta_i^7)$ *as* $\theta_i \to 0$.

5. Asymptotic inference

This section develops the asymptotic theory for the MLE and then gives asymptotically valid standard error and confidence region under the proposed model.

5.1. Asymptotic theory

The proposed method deals with independent but not identically distributed (i.n.i.d.) samples due to the heterogeneity of covariance matrices across $i = 1, 2, \cdots, n$. This implies that the well-known asymptotic theory for MLEs under independent and identically distributed (i.i.d.) samples is not suitable for our setting.

The classical paper of Bradly and Gart [26] studied the asymptotic properties of MLEs when samples are i.n.i.d.. However, as pointed out by Emura et al. [27], the regularity conditions of Bradly and Gart [26] are not reliable as the proof of their theory is far from complete and relies on a less known theorem (Khintchine's theorem). Here, we reform the regularity conditions of Bradly and Gart [26] to be adapted to more common tools on probability theory, such as the weak law of large numbers (WLLN) and the Lindeberg-Feller central limit theorem (CLT) for i.n.i.d. random variables. The WLLN for i.n.i.d. random variables from Theorem 1.14 in Shao [28] is stated as follows:

The WLLN *Let* $Y_1, Y_2, \cdots, Y_n$ *be independent random variables with* $E\{|Y_i|\} < \infty$ *for* $i = 1, 2, \cdots, n$. *If there exists a constant* $p \in [1, 2]$ *such that* $n^{-p}\Sigma_{i=1}^{n} E\{|Y_i|^p\} \to 0$. *Then*

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i - E(Y_i)\} \xrightarrow{p} 0, \quad as\ n \to \infty,$$

where '$\xrightarrow{p}$' *denotes convergence in probability.*

Let $\mathbf{1}\{\cdot\}$ be the indicator function and Cov($\mathbf{D}$) be the covariance matrix of a random vector $\mathbf{D}$. The Lindeberg-Feller multivariate CLT from Proposition 2.27 in van der Vaart [29] is stated as follows:

The Lindeberg-Feller CLT *Let* $\mathbf{D}_{n,1}, \mathbf{D}_{n,2}, \cdots, \mathbf{D}_{n,n}$ *be independent 2-dimensional random vectors with finite second moments such that*

$$\sum_{i=1}^{n} E[||\mathbf{D}_{n,i} - E(\mathbf{D}_{n,i})||^2 \mathbf{1}\{||\mathbf{D}_{n,i} - E(\mathbf{D}_{n,i})|| > \varepsilon\}] \to 0, \quad as\ n \to \infty \tag{6}$$

for all $\varepsilon > 0$, *and* $\Sigma_{i=1}^{n}\text{Cov}(\mathbf{D}_{n,i}) \to \Sigma$ *for a positive definite matrix* Σ. *Here* $||\cdot||$ *denotes the Euclidean norm for 2-dimensional vectors. Then,*

$$\sum_{i=1}^{n}\{\mathbf{D}_{n,i} - E(\mathbf{D}_{n,i})\} \xrightarrow{d} \text{BVN}(\mathbf{0}, \Sigma), \qquad \text{as } n \to \infty,$$

where '$\xrightarrow{d}$' *denotes convergence in distribution.*

Equation (6) is known as the Lindeberg condition. We state the regularity conditions.

Assumption (A): The parameter space Θ is open and contains the true parameter point $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)^{\mathrm{T}}$.

For instance, $\Theta = \mathbb{R}^2$ and $\Theta = (0,\infty)^2$ satisfy Assumption (A), but $\Theta = [0,\infty)^2$ does not. In general, Assumption (A) holds unless unusual constraints on Θ are imposed.

Assumption (B): There exists a 2×2 positive definite matrix I such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I_i = I,$$

where I is called the large sample Fisher information matrix.

Assumption (C): For $j, k, \ell = 1, 2$, there exist finite constants $w_{jk\ell}$ and m_{jk} such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\mu}}\left\{\left|\frac{\partial^3 \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial\mu_j\partial\mu_k\partial\mu_\ell}\right|\right\} = w_{jk\ell},$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\mu}}\left[\left\{\frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial\mu_j}\right\}^2\left\{\frac{\partial \log f_{i,\boldsymbol{\mu}}(\mathbf{Y}_i)}{\partial\mu_k}\right\}^2\right] = m_{jk}.$$

Assumption (D): For $m = 1, 2, 3$, there exist constants $s_{j,m}, j = 1, 2$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_{i1}^{2m}} = s_{1,m} < \infty, \quad \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_{i2}^{2m}} = s_{2,m} < \infty.$$

Assumptions (B)–(D) hold if $(\sigma_{i1}^2, \sigma_{i2}^2, \theta_i) = (\sigma_1^2, \sigma_2^2, \theta)$ for all i's. More generally, they hold if $(\sigma_{i1}^2, \sigma_{i2}^2, \theta_i)$ are stable with respect to i. The expectations in Assumption (C) can be written as double integrals which are independent of $\boldsymbol{\mu}$ as in Theorem 4.1. Assumption (D) ensures the stability of variances among the studies. Assumptions (B) – (D) are required to verify the conditions in the WLLN and the Lindeberg-Feller CLT.

We examine Assumption (D) through a concrete example. Suppose that σ_{i1}^2's are independently sampled from $X|u < X < v$ which follows a truncated gamma distribution with a shape parameter $\alpha > 0$, a scale parameter $\beta > 0$, and a truncation interval $[u, v]$, where $0 < u < v \leq \infty$. By the strong law of large number for i.i.d. samples, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_{i1}^{2m}} \to E(X^{-m}|u < X < v) < \infty, \qquad \text{as } n \to \infty, m = 1, 2, 3.$$

The truncation ensures the existence of the expectation $E(X^{-m}|u < X < v)$. Hence Assumption (D) holds by defining $s_{1,m} \equiv E(X^{-m}|u < X < v)$. If $\alpha > m$, we have an explicit formula

$$s_{1,m} = E(X^{-m}|u < X < v) = \frac{\Gamma(\alpha - m)\{P(\alpha - m, v/\beta) - P(\alpha - m, u/\beta)\}}{\Gamma(\alpha)\beta^m\{P(\alpha, v/\beta) - P(\alpha, u/\beta)\}},$$

where

$$P(z,t) = \frac{1}{\Gamma(z)}\int_0^t x^{z-1}e^{-x}dx, \Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx.$$

On the contrary, if there is no truncation, then the expectation $E(X^{-m})$ does not exist when $\alpha \le m$. This implies that Assumption (D) does not hold. Our simulation studies of Section 6 shall use the setting of $\alpha = 1/2$, $\beta = 1/2$, and $[u, v] = [0.009, 0.6]$.

Theorem 5.1: *If Assumptions (A) - (D) hold, then*

(a) *Existence and consistency: With probability tending to one, there exists the MLE* $\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}} = (\hat{\mu}_{1,n}^{\mathrm{MLE}}, \hat{\mu}_{2,n}^{\mathrm{MLE}})^{\mathrm{T}}$ *such that* $\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}} \xrightarrow{p} \boldsymbol{\mu}^0$, *as* $n \to \infty$.
(b) *Asymptotic normality:* $\sqrt{n}(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}} - \boldsymbol{\mu}^0) \xrightarrow{d} \mathrm{BVN}(0, I^{-1})$, *as* $n \to \infty$.

Theorem 5.1 is proven by approximating $\ell_n(\boldsymbol{\mu}) - \ell_n(\boldsymbol{\mu}^0)$ and $\sqrt{n}(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}} - \boldsymbol{\mu}^0)$ to the sum of i.n.i.d. random variables and applying the WLLN and CLT for i.n.i.d. samples to them. The proof of Theorem 5.1 mainly follows from Theorem 6.5.1 in Lehmann and Casella [30] and is given in Supplementary Material.

5.2. Standard error and confidence interval

We provide three different approaches to obtain the standard error (SE) of $g(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}})$ and confidence interval (CI) of $g(\boldsymbol{\mu})$, where $g : \mathrm{R}^2 \mapsto \mathrm{R}$ is a differentiable function.

(i) Using the exact Fisher information

Under Assumption (B) and for large n, we approximate the large sample Fisher information matrix by $I \approx \Sigma_{i=1}^n I_i/n$. By Theorem 5.1 (b) and the delta method, the SE of $g(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}})$ is

$$\mathrm{SE}^{\mathrm{Exact}}\{g(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}})\} = \sqrt{\left\{\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right\}^{\mathrm{T}}\left\{\sum_{i=1}^n I_i\right\}^{-1}\left\{\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right\}\Bigg|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}}}.$$

(ii) Using the approximate Fisher information

The second approach is an application of Theorem 4.5. When n is large, we obtain another approximation of the large sample Fisher information matrix by $I \approx \Sigma_{i=1}^n I_i/n \approx$

$\Sigma_{i=1}^{n}\tilde{I}_i/n$, where $\tilde{I}_i$ is the approximate Fisher information matrix in Theorem 4.5. Thus,

$$\tilde{I}_i = \begin{bmatrix} \dfrac{0.012 \times \theta_i^4 + 0.1224 \times \theta_i^2 + 1}{\sigma_{i1}^2} & -\dfrac{0.0032 \times \theta_i^5 + 0.0148 \times \theta_i^3 + \pi^{-1}\theta_i}{\sigma_{i1}\sigma_{i2}} \\ -\dfrac{0.0032 \times \theta_i^5 + 0.0148 \times \theta_i^3 + \pi^{-1}\theta_i}{\sigma_{i1}\sigma_{i2}} & \dfrac{0.012 \times \theta_i^4 + 0.1224 \times \theta_i^2 + 1}{\sigma_{i2}^2} \end{bmatrix}.$$

Accordingly, the SE of $g(\hat{\boldsymbol{\mu}}_n^{\text{MLE}})$ is

$$\text{SE}^{\text{Approx}}\{g(\hat{\boldsymbol{\mu}}_n^{\text{MLE}})\} = \sqrt{\left\{\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right\}^{\text{T}} \left\{\sum_{i=1}^{n} \tilde{I}_i\right\}^{-1} \left\{\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right\}\Bigg|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_n^{\text{MLE}}}}.$$

(iii) Using the observed Fisher information

The last approach is based on Equation (4). When n is large, we approximate the large sample Fisher information matrix by the observed information matrix

$$I \approx \frac{1}{n}\sum_{i=1}^{n} I_i \approx \frac{1}{n}\sum_{i=1}^{n} \hat{I}_i(\hat{\boldsymbol{\mu}}_n^{\text{MLE}}) \equiv -\frac{1}{n}\frac{\partial^2 \ell_n(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}\partial \boldsymbol{\mu}^{\text{T}}}\Bigg|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_n^{\text{MLE}}}.$$

This quantity is available from the NR algorithm. The SE of $g(\hat{\boldsymbol{\mu}}_n^{\text{MLE}})$ is

$$\text{SE}^{\text{Obs}}\{g(\hat{\boldsymbol{\mu}}_n^{\text{MLE}})\} = \sqrt{\left\{\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right\}^{\text{T}} \left\{-\frac{\partial^2 \ell_n(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}\partial \boldsymbol{\mu}^{\text{T}}}\right\}^{-1} \left\{\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right\}\Bigg|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_n^{\text{MLE}}}}.$$

The approach (i) is computationally most difficult as it involves double integrals. However, it has a good theoretical support in the light of Theorem 5.1. The approach (ii) is substantially easier to compute than the approach (i), but relies on the non-asymptotic approximation. Both the approaches (i) and (ii) do not involve any unknown quantity if the function $\partial g(\boldsymbol{\mu})/\partial \boldsymbol{\mu}$ does not depend on $\boldsymbol{\mu}$. For instance, if $g(\boldsymbol{\mu}) = \mu_2 - \mu_1$, then $\{\partial g(\boldsymbol{\mu})/\partial \boldsymbol{\mu}\}^{\text{T}} = (-1, 1)^{\text{T}}$ is independent of $\boldsymbol{\mu}$. The approach (iii) is also easy to compute, but involves estimation of $\boldsymbol{\mu}$. This seems unnatural since the Fisher information matrix does not depend on $\boldsymbol{\mu}$ (Theorem 4.1). In all the three approaches, the 95% confidence interval is $g(\hat{\boldsymbol{\mu}}_n^{\text{MLE}}) \pm 1.96 \times \text{SE}\{g(\hat{\boldsymbol{\mu}}_n^{\text{MLE}})\}$.

5.3. *Confidence ellipsoid*

This section considers the confidence ellipsoid (CE) for $\boldsymbol{\mu}$. By Theorem 5.1 (b),

$$(\hat{\boldsymbol{\mu}}_n^{\text{MLE}} - \boldsymbol{\mu})^{\text{T}} n\hat{I}(\hat{\boldsymbol{\mu}}_n^{\text{MLE}} - \boldsymbol{\mu}) \xrightarrow{d} \chi^2_{\text{df}=2}, \quad as\, n \to \infty,$$

where $\hat{I}$ is a consistent estimator of I. Let $\chi^2_{\text{df}=2,0.95}$ be the 0.95 percentile of the chi-squared distribution with two degrees of freedom. We construct a 95% CE for $\boldsymbol{\mu}$ based on the three different approaches of estimating I.

(i) Using the exact Fisher information

$$\mathrm{CE}^{\mathrm{Exact}}=\left\{\boldsymbol{\mu}:(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}-\boldsymbol{\mu})^{\mathrm{T}}\sum_{i=1}^{n}I_i(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}-\boldsymbol{\mu})\le\chi^2_{\mathrm{df}=2,0.95}\right\}.$$

(ii) Using the approximate Fisher information

$$\mathrm{CE}^{\mathrm{Approx}}=\left\{\boldsymbol{\mu}:(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}-\boldsymbol{\mu})^{\mathrm{T}}\sum_{i=1}^{n}\tilde{I}_i(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}-\boldsymbol{\mu})\le\chi^2_{\mathrm{df}=2,0.95}\right\}.$$

(iii) Using the observed Fisher information

$$\mathrm{CE}^{\mathrm{Obs}}=\left\{\boldsymbol{\mu}:(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}-\boldsymbol{\mu})^{\mathrm{T}}\left(-\frac{\partial^2\ell_n(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\mathrm{T}}}\bigg|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}}\right)(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}}-\boldsymbol{\mu})\le\chi^2_{\mathrm{df}=2,0.95}\right\}.$$

5.4. Percentage study weight

Under the bivariate FGM model, we consider the percentage study weight [31] which shows the relative contribution of each individual study for estimating the common mean vector. According to Riley et al. [31], the percentage study weight of study i for $\hat{\mu}_j^{\mathrm{BN}}$ under the bivariate normal model is

$$100\%\times\mathbf{W}_{i,jj}/\Sigma_{i=1}^{n}\mathbf{W}_{i,jj},\mathbf{W}_i=(\Sigma_{i=1}^{n}\mathbf{C}_i^{-1})^{-1}\mathbf{C}_i^{-1}(\Sigma_{i=1}^{n}\mathbf{C}_i^{-1})^{-1},\ \ j=1,2,i=1,2,\cdots,n.$$

Analogously, we derive the percentage study weight of study i for $\hat{\mu}_j^{\mathrm{MLE}}$ under the bivariate FGM model and it is

$$100\%\times\mathbf{H}_{i,jj}/\Sigma_{i=1}^{n}\mathbf{H}_{i,jj},\mathbf{H}_i=(\Sigma_{i=1}^{n}I_i)^{-1}I_i(\Sigma_{i=1}^{n}I_i)^{-1},\ j=1,2,i=1,2,\cdots,n.$$

Our derivations are based on the asymptotic approximation (Theorem 5.1)

$$\mathrm{var}(\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}})\approx(nI)^{-1}\approx(\Sigma_{i=1}^{n}I_i)^{-1}.$$

6. Simulation

We conduct Monte Carlo simulations to examine the performance of the proposed method.

We generate data $\mathbf{Y}_i=(Y_{i1},Y_{i2})^{\mathrm{T}}$ following the FGM model in Equation (2), for $i=1,2,\cdots,n$. Without loss of generality, the mean vector is set to be $\boldsymbol{\mu}=(0,0)^{\mathrm{T}}$. To set the known variances σ_{i1}^2 and σ_{i2}^2, we mimic the simulation setting of Kontopantelis and Reeves [32]. That is, $\sigma_{i1}^2,\sigma_{i2}^2\sim\chi^2_{\mathrm{df}=1}/4$, restricted in the interval $[0.009,0.6]$. This leads to $E[\sigma_{i1}^2]=E[\sigma_{i2}^2]=0.173$. We generate the dependence parameter θ_i from the beta distribution Beta$(27,3)$, Beta$(42.5,42.5)$, or Beta$(3,27)$, corresponding to stronger ($E[\theta_i]=0.9$), medium ($E[\theta_i]=0.5$), or weaker ($E[\theta_i]=0.1$) dependence, respectively.

Based on the data, we compute $\hat{\mu}_{1,n}^{\mathrm{MLE}}$ and $\hat{\mu}_{2,n}^{\mathrm{MLE}}-\hat{\mu}_{1,n}^{\mathrm{MLE}}$ by using the NR Algorithm, and count the number of iterations to assess the convergence speed. We calculate the three SEs ($\mathrm{SE}^{\mathrm{Exact}}$, $\mathrm{SE}^{\mathrm{Approx}}$, and $\mathrm{SE}^{\mathrm{Obs}}$) and evaluate the coverage probability (CP) of the three 95% CIs. In addition, we also examine the square error loss defined as

Table 1. Simulation results on $\hat{\mu}_{1,n}^{\text{MLE}}$ and $\hat{\mu}_{2,n}^{\text{MLE}} - \hat{\mu}_{1,n}^{\text{MLE}}$ based on 1,000 repetitions.

		$\hat{\mu}_{1,n}^{\text{MLE}}$							$\hat{\mu}_{2,n}^{\text{MLE}} - \hat{\mu}_{1,n}^{\text{MLE}}$						
			Exact		Approx		Obs			Exact		Approx		Obs	
Parameters	n	SD	SE	CP	SE	CP	SE	CP	SD	SE	CP	SE	CP	SE	CP
$E[\theta_i] = 0.9$	5	0.123	0.119	0.954	0.119	0.954	0.120	0.957	0.154	0.171	0.950	0.171	0.951	0.172	0.949
	10	0.079	0.077	0.936	0.077	0.937	0.077	0.943	0.097	0.111	0.956	0.111	0.956	0.111	0.953
	15	0.063	0.062	0.953	0.062	0.953	0.062	0.953	0.078	0.088	0.956	0.088	0.956	0.088	0.955
$E[\theta_i] = 0.5$	5	0.124	0.121	0.958	0.121	0.958	0.121	0.961	0.163	0.173	0.959	0.173	0.959	0.174	0.960
	10	0.086	0.080	0.949	0.080	0.949	0.080	0.947	0.109	0.113	0.947	0.113	0.947	0.113	0.948
	15	0.061	0.063	0.963	0.063	0.963	0.063	0.960	0.085	0.090	0.948	0.090	0.948	0.090	0.947
$E[\theta_i] = 0.1$	5	0.126	0.122	0.950	0.122	0.950	0.122	0.950	0.174	0.176	0.964	0.176	0.964	0.176	0.964
	10	0.080	0.079	0.943	0.079	0.943	0.079	0.943	0.111	0.114	0.960	0.114	0.960	0.114	0.961
	15	0.064	0.064	0.953	0.064	0.953	0.064	0.953	0.089	0.090	0.959	0.090	0.959	0.090	0.960

SD = standard deviation, SE = standard error, CP = coverage probability of the 95% CI, Exact = exact Fisher information, Approx = approximate Fisher information, Obs = observed Fisher information.

Table 2. Simulation results on $\hat{\boldsymbol{\mu}}_n^{\text{MLE}}$ based on 1,000 repetitions.

				Exact	Approx	Obs
Parameters	n	No. of iterations	Square error loss	CP	CP	CP
$E[\theta_i] = 0.9$	5	3.0	0.0304	0.952	0.952	0.949
	10	2.9	0.0128	0.949	0.949	0.951
	15	2.8	0.0077	0.949	0.950	0.951
$E[\theta_i] = 0.5$	5	2.8	0.0311	0.959	0.959	0.959
	10	2.6	0.0138	0.953	0.953	0.951
	15	2.6	0.0083	0.950	0.950	0.951
$E[\theta_i] = 0.1$	5	2.0	0.0317	0.954	0.954	0.952
	10	2.0	0.0132	0.960	0.960	0.959
	15	2.0	0.0081	0.951	0.951	0.952

CP = coverage probability of the 95% CE, Exact = exact Fisher information, Approx = approximate Fisher information, Obs = observed Fisher information.

$L(\hat{\boldsymbol{\mu}}_n^{\text{MLE}}, \boldsymbol{\mu}) = (\hat{\boldsymbol{\mu}}_n^{\text{MLE}} - \boldsymbol{\mu})^{\text{T}}(\hat{\boldsymbol{\mu}}_n^{\text{MLE}} - \boldsymbol{\mu})$ and evaluate the CP of the 95% CEs (CE^{Exact}, $\text{CE}^{\text{Approx}}$, and CE^{Obs}). Our simulations are based on $n \in \{5, 10, 15\}$ and 1,000 repetitions.

Table 1 displays the performance of $\hat{\mu}_{1,n}^{\text{MLE}}$ and $\hat{\mu}_{2,n}^{\text{MLE}} - \hat{\mu}_{1,n}^{\text{MLE}}$. The standard deviation (SD) decreases when the sample size increases from $n = 5$ to $n = 15$. The SEs are all very close to the SDs. The close proximity of $\text{SE}^{\text{Approx}}$ to SE^{Exact} indicates that the proposed approximation is fairly accurate. The CPs of the 95% CIs are reasonably close to the 95%.

Table 2 shows the average number of iterations in the NR algorithm. It reveals that the NR Algorithm converges very quickly (2–3 iterations on average). As expected, stronger dependence ($E[\theta_i] = 0.9$) requires more iterations than weaker dependence ($E[\theta_i] = 0.1$). Table 2 also shows that the square error loss of $\hat{\boldsymbol{\mu}}_n^{\text{MLE}}$ decreases when the sample size increases from $n = 5$ to $n = 15$. The CPs of the 95% CEs are all close to the nominal level.

Overall, the proposed methods yield sound performance on estimates, SEs, square error loss, and CPs. This implies that the asymptotic inference of Section 5 works fairly well.

7. Data analysis

We analyse the entrance examination data for entering Graduate Institute of Statistics, National Central University (NCU), Taiwan. The data consist of mathematics and

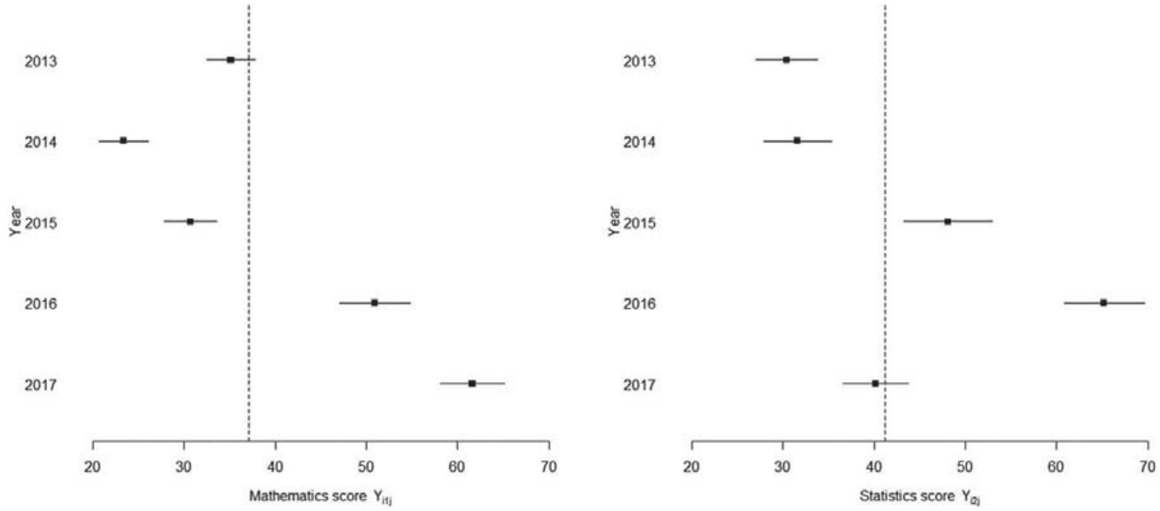

Figure 3. The mean scores of mathematics (Y_{i1}) and statistics (Y_{i2}) based on the entrance examination data. The horizontal line denotes $Y_{ij} \pm 1.96\sigma_{ij}^2$ for $j = 1, 2$. The vertical line denotes the estimator of the common mean.

statistics scores of 848 students who took written exams from 2013 to 2017. Thus, we let $i = 1, 2, \cdots, 5$ corresponding to 2013, 2014, ..., 2017. The possible range of score is from 0 to 100 for both subjects. The number of applicants monotonically increase from 2013 ($n_1 = 148$) to 2017 ($n_5 = 198$). The data are the official records from Admission Division of NCU.

For each academic year, we compute the mean scores of mathematics (Y_{i1}) and statistics (Y_{i2}), and their covariance matrix ($\mathbf{C}_i$) by using the individual scores. Specifically, we let

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} = \frac{1}{n_i} \sum_{j=1}^{n_i} \begin{bmatrix} Y_{i1j} \\ Y_{i2j} \end{bmatrix},$$

$$\mathbf{C}_i = \begin{bmatrix} \sigma_{i1}^2 & \rho_i \sigma_{i1} \sigma_{i2} \\ \rho_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2 \end{bmatrix} = \frac{1}{n_i(n_i - 1)} \sum_{j=1}^{n_i} \begin{bmatrix} Y_{i1j} - Y_{i1} \\ Y_{i2j} - Y_{i2} \end{bmatrix} \begin{bmatrix} Y_{i1j} - Y_{i1} \\ Y_{i2j} - Y_{i2} \end{bmatrix}^{\mathrm{T}},$$

where $(Y_{i1j}, Y_{i2j})^{\mathrm{T}}$ is a vector of scores for student j and n_i is the number of students in year i. The values of $\mathbf{Y}_i$'s are shown in Figure 3 and summarized in Table 3. The subsequent meta-analysis is solely based $\mathbf{Y}_i$'s and $\mathbf{C}_i$'s.

We fit the data to the bivariate FGM model in Equation (2). First, the dependence parameter θ_i, $i = 1, 2, \cdots, 5$ is estimated by the relationship $\rho_i = \theta_i/\pi$ with some boundary corrections to meet $\theta_i \in [-1, 1]$ [22]. Thus, we let $\theta_i = \min[1, \max\{-1, \pi\rho_i\}]$ for $i = 1, 2, \cdots, 5$. Accordingly, we obtain $\theta_i = 1$, $i = 1, 2, \cdots, 5$. Then, we use the NR Algorithm to obtain the MLE $\hat{\boldsymbol{\mu}}_n^{\mathrm{MLE}} = (\hat{\mu}_{1,n}^{\mathrm{MLE}}, \hat{\mu}_{2,n}^{\mathrm{MLE}}) = (37.16, 41.17)$. The fitted results are summarized in Table 4 and Figure 4. The three different ways of calculating CIs give very similar results (Table 4). This agrees with the simulation results.

Our analysis reveals that the mean scores of mathematics and statistics are significantly below 50. Considering that the maximum score is 100 for both subjects, majority of students performed poorly on the exams. The 95% CE for $\boldsymbol{\mu}$ confirms this observation as the CE is far away from $\boldsymbol{\mu} = (50, 50)$ (Figure 4). The mean score of statistics ($\hat{\mu}_{2,n}^{\mathrm{MLE}} = 41.17$) is

Table 3. Summary statistics for the entrance exam data.

i	Year	Number of students (n_i)	Mathematics score (Y_{i1})	Statistics score (Y_{i2})	Covariance matrix (C_i) Math Stat	ρ_i	θ_i
1	2013	148	35.17	30.41	$\begin{bmatrix} 1.77 & 0.89 \\ 0.89 & 2.99 \end{bmatrix}$	0.38	1
2	2014	155	23.43	31.63	$\begin{bmatrix} 1.89 & 1.76 \\ 1.76 & 3.61 \end{bmatrix}$	0.67	1
3	2015	167	30.74	48.11	$\begin{bmatrix} 2.15 & 2.12 \\ 2.12 & 6.13 \end{bmatrix}$	0.58	1
4	2016	180	50.91	65.22	$\begin{bmatrix} 3.87 & 2.91 \\ 2.91 & 5.02 \end{bmatrix}$	0.66	1
5	2017	198	61.62	40.22	$\begin{bmatrix} 3.17 & 2.10 \\ 2.10 & 3.29 \end{bmatrix}$	0.65	1

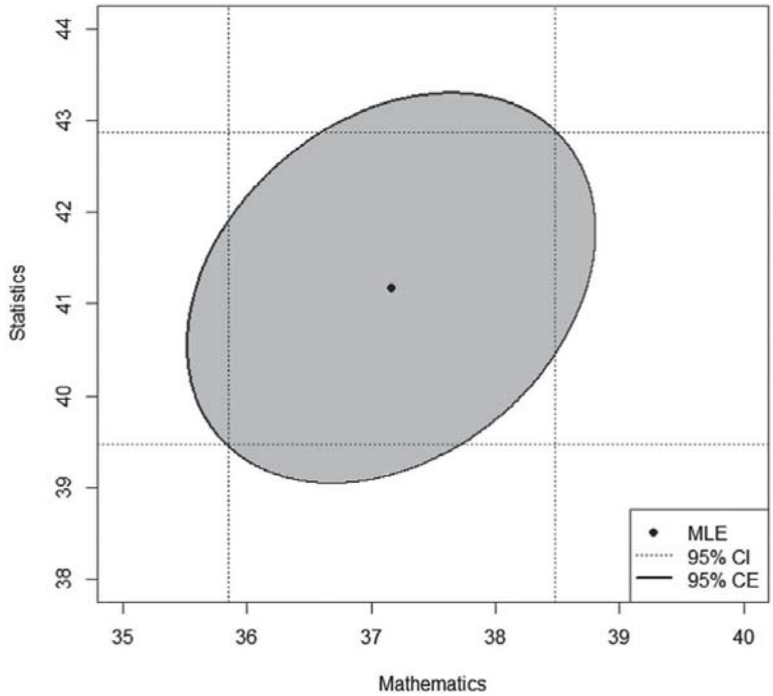

Figure 4. The MLE and the 95% CI and CE for the common means based on the exact Fisher information matrix.

slightly higher than the mean score of mathematics ($\hat{\mu}_{1,n}^{\mathrm{MLE}} = 37.16$). This may be because some mathematics problems are too difficult for students from the schools of humanities.

We compare the MLEs between the FGM model and the bivariate normal model. Under the bivariate normal model, the MLE becomes $\hat{\boldsymbol{\mu}}_n^{\mathrm{BN}} = (\Sigma_{i=1}^n \mathbf{C}_i^{-1})^{-1} \Sigma_{i=1}^n \mathbf{C}_i^{-1} \mathbf{Y}_i = (35.83, 38.64)$. These mean scores appear unnaturally low. Indeed, a worse fit of the bivariate normal model is indicated by its inferior log-likelihood value (Table 5). This is mainly caused by the poor fit for the scores in 2017, which contains the largest number of students.

The large value of the log-likelihood function under the bivariate FGM copula model is partly due to the boundary correction $\theta_i = \min[1, \max\{-1, \pi\rho_i\}]$. Accordingly, the fitted values for the correlation are decreased by the boundary correction. However, the log-likelihood value under the bivariate normal model will increase if we also decrease its

Table 4. Estimation results for the entrance exam data.

Model		Mathematics	Statistics
FGM	Estimate	37.16	41.17
	95% CI (Exact)	(35.85, 38.48)	(39.48, 42.87)
	(Approx)	(35.85, 38.48)	(39.47, 42.88)
	(Obs)	(35.85, 38.47)	(39.65, 42.70)
Bivariate Normal	Estimate	35.83	38.64
	95% CI	(34.51, 37.16)	(36.94, 40.34)

95% CI (Approx) is computed by setting $\theta_i = 1 - 10^{-10}, i = 1, \cdots, 5$.

Table 5. The individual log-likelihood values calculated for the entrance exam data.

	Model	2013	2014	2015	2016	2017	Total
Log-likelihood value	FGM	−22.52	−64.60	−21.91	−84.63	−98.14	−291.80
	Bivariate normal	−14.97	−48.33	−34.96	−73.47	−170.92	−342.65

Log-likelihood (Total):

$$\ell_5^{\rm FGM}(\hat{\boldsymbol{\mu}}_5^{\rm MLE}) = \sum_{i=1}^{5} \log Q_i(\mathbf{Y}_i; \hat{\boldsymbol{\mu}}_5^{\rm MLE}) - \sum_{j=1}^{2}\sum_{i=1}^{5} \frac{(Y_{ij} - \hat{\mu}_{j,5}^{\rm MLE})^2}{2\sigma_{ij}^2} - 5\log(2\pi) - \sum_{j=1}^{2}\sum_{i=1}^{5} \log\sigma_{ij} = -291.80,$$

$$\ell_5^{\rm BN}(\hat{\boldsymbol{\mu}}_5^{\rm BN}) = \sum_{i=1}^{5} \frac{\rho_i(Y_{i1} - \hat{\mu}_{1,5}^{\rm BN})(Y_{i2} - \hat{\mu}_{2,5}^{\rm BN})}{(1-\rho_i^2)\sigma_{i1}\sigma_{i2}} - \sum_{j=1}^{2}\sum_{i=1}^{5} \frac{(Y_{ij} - \hat{\mu}_{j,5}^{\rm BN})^2}{2(1-\rho_i^2)\sigma_{ij}^2} - 5\log(2\pi) - \sum_{j=1}^{2}\sum_{i=1}^{5} \log\sigma_{ij}$$
$$- \frac{1}{2}\sum_{i=1}^{5}\log(1-\rho_i^2) = -342.65.$$

Table 6. Percentage study weights based on the entrance exam data.

	Subject	Model	2013 ($n_1 = 148$)	2014 ($n_2 = 155$)	2015 ($n_3 = 167$)	2016 ($n_4 = 180$)	2017 ($n_5 = 198$)
Percentage study weights	Mathematics	FGM	26.4%	24.8%	21.8%	12.1%	14.9%
		Bivariate normal	26.9%	24.9%	21.8%	11.8%	14.6%
	Statistics	FGM	26.1%	21.6%	12.8%	15.6%	23.9%
		Bivariate normal	26.5%	20.9%	12.9%	15.5%	24.2%

correlation. This phenomenon may relate to our ignorance of the between-study variance (heterogeneity) of the mean scores. One potential way to clarify this problem is to consider a random-effects model which is our future work.

Finally, we report the percentage study weight that is the contribution of each year for estimating the common mean score (Section 5.4). Table 6 reveals that the percentage study weights are similar between the bivariate FGM copula and bivariate normal models. The largest percentage corresponds to the year 2013 ($i = 1$) which has the smallest within-study variance for both two subjects ($\mathbf{C}_{1,11} = 1.77$ and $\mathbf{C}_{1,22} = 2.99$).

8. Summary and discussions

This paper proposes a likelihood-based estimation method for a common mean vector in bivariate meta-analysis under the FGM copula model. We have established key theoretical results behind the estimation method, including the expressions of the Fisher information matrix and the validity of the asymptotic theory under the i.n.i.d samples. We also provide three different approaches to construct a confidence interval (region) for the common mean. We show by simulations that our estimation method is nearly unbiased and provides fairly accurate coverage probability of the confidence interval. Our real data analysis demonstrates the case where the bivariate FGM model fits better than the bivariate normal model.

The proposed method based on the bivariate FGM model is not simply an alternative to the bivariate normal model. Even if the bivariate normal model is the main model, a sensitivity analysis based on the bivariate FGM model would be useful. Indeed, our data real analysis revealed that the estimates are remarkably different between two models. Such results typically imply that at least one of the two models is not adequate to the data. In such a circumstance, we suggest choosing a better model by comparing their likelihood values.

In the literature, many authors considered different ways to generalize the FGM copula [18,33–37]. There are two merits to consider these generalized FGM copulas. First, the generalized FGM copulas may extend the narrow range of correlation in the FGM copula. Second, the generalized FGM copulas can introduce the asymmetric structure and may be useful to study directional dependence [18]. Bivariate meta-analysis under the generalized FGM copula is a topic for future research.

We shall consider the extension of the current FGM model to allow survival data. In meta-analysis of bivariate survival data, it is common to apply copula models with the Weibull margins (e.g., [3]), spline margins [6,38], or nonparametric margins [39]. The Clayton copula has been the most common choice among other copulas. This is because the Clayton copula has simple derivative functions [40], making it suitable for deriving a likelihood function even in the presence of censoring in meta-analysis [6,38]. However, the Fisher information of the Clayton copula is not very simple even under complete data [40]. Our recent paper shows that the generalized FGM copula proposed by Bairamov and Kotz [35] gives simple expressions for a likelihood-based analysis with bivariate competing risks data in a single study setting [41].

Acknowledgements

The authors thank Editor, Associate Editor, and two anonymous reviewers for their detailed and insightful comments and suggestions that greatly improved the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was presented in the Japanese Joint Statistical Meeting 2018, Tokyo, Japan. Our work is supported by the government of Taiwan (MOST 105-2118-M-008-001-MY2; 107-2118-M-008-003-MY3) (Ministry of Science and Technology, Taiwan).

References

[1] Gleser LJ, Olkin L. Stochastically dependent effect sizes. In the Handbook of research Synthesis. New York: Russel Sage Foundation; 1994.

[2] Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. J R Statist Soc Ser A. 2009;172:789–811.
[3] Burzykowski T, Molenberghs G, Buyse M, et al. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. Appl Stat. 2001;50:405–422.
[4] Rotolo F, Paoletti X, Michiels S. Surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. Comput Methods Programs Biomed. 2018;155:189–198.
[5] Emura T, Nakatochi M, Matsui S, et al. Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: meta-analysis with a joint model. Stat Methods Med Res. 2018;27(9):2842–2858.
[6] Emura T, Matsui S, Rondeau V. Survival analysis with Correlated Endpoints, joint Frailty-copula models, JSS research Series in statistics. Singapore: Springer; 2019.
[7] Berkey CS, Hoaglin DC, Antczak-bouckoms A, et al. Meta-analysis of multiple outcomes by regression with random effects. Stat Med. 1998;17:2537–2550.
[8] Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. Stat Method Med Res. 2013;22:133–158.
[9] Morgenstern D. Einfache Beispiele zweidimensionaler Verteilungen. Mitteilungsblatt für Mathematishe Statistik. 1956;8:34–235.
[10] Eyraud H. (1936). Les principes de la mesure des correlations, Ann Univ Lyon III Ser Sect A1, 30-47.
[11] Farlie DJG. The performance of some correlation coefficients for a general bivariate distribution. Biometrika. 1960;47:307–323.
[12] Gumbel EJ. Bivariate exponential distributions. J Amer Statist Assoc. 1960;55:698–707.
[13] Huang JS, Kotz S. Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions. Biometrika. 1984;71:633–636.
[14] Baker R. An order-statistics-based method for constructing multivariate distributions with fixed marginals. J Multivariate Anal. 2008;99:2312–2327.
[15] Bairamov I, Bayramoglu K. From the Huang-Kotz FGM distribution to Baker's bivariate distribution. J Multivariate Anal. 2013;113:106–115.
[16] Huang JS, Dou X, Kuriki S, et al. Dependence structure of bivariate order statistics with applications to Bayramoglu's distributions. J Multivariate Anal. 2013;114:201–208.
[17] Genest C, Favre AC. Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng. 2007;12:347–368.
[18] Kim JM, Jung YS, Soderberg T. Directional dependence of genes using survival truncated FGM type modification copulas. Commun Stat – Simul Comput. 2009;38:1470–1484.
[19] Martinez EZ, Achcar JA. Bayesian bivariate generalized Lindley model for survival data with a cure fraction. Comput Method Programs Biomed. 2014;117:145–157.
[20] Shinozaki N. A note on estimating the common mean of k normal distributions and the stein problem. Commun Stat – Theory Methods. 1978;7:1421–1432.
[21] Lehmann EL. Elements of large-sample theory. New York: Springer; 2004.
[22] Genest C, Nešlehová J. Copulas and copula models, Encyclopedia of Environmetrics. 2nd ed Chichester: Wiley; 2012.
[23] Schucany WR, Parr WC, Boyer JE. Correlation structure in Farlie-Gumbel-Morgenstern distributions. Biometrika. 1978;65:650–653.
[24] Nelsen RB. An Introduction to copulas. (2nd ed.) New York: Springer; 2006.
[25] Casella G, Berger RL. Statistical inference. (2nd ed.) Australia: Duxbury Press; 2002.
[26] Bradley RA, Gart JJ. The asymptotic properties of ML estimators when sampling from associated populations. Biometrika. 1962;49:205–214.
[27] Emura T, Hu YH, Konno Y. Asymptotic inference for maximum likelihood estimators under the special exponential family with double-truncation. Stat Pap. 2017a;58:877–909.
[28] Shao J. Mathematical statistics. New York: Springer; 2003.
[29] van der Vaart AW. Asymptotic statistics. Cambridge: Cambridge University Press; 1998.
[30] Lehmann EL, Casella G. Theory of point estimation. (2nd ed.) New York: Springer; 1998.

[31] Riley RD, Ensor J, Jackson D, et al. 2017. Deriving percentage study weights in multi-parameter meta-analysis models: with applications to meta-regression, network meta-analysis and one-stage individual participant data models, Stat Method Med Res. doi:10.1177/0962280216688033.
[32] Kontopantelis E, Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. Stat Method Med Res. 2012;21:409–426.
[33] Huang JS, Kotz S. Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. Metrika. 1999;49:135–145.
[34] Bairamov I, Kotz S, Bekci M. New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics. J Appl Stat. 2001;28:521–536.
[35] Bairamov I, Kotz S. Dependence structure and symmetry of Huang-Kotz FGM distributions and their extensions. Metrika. 2002;56:55–72.
[36] Amblard C, Girard S. A new extension of bivariate FGM copulas. Metrika. 2009;70:1–17.
[37] Hürlimann W. A comprehensive extension of the FGM copula. Stat Pap. 2017;58:373–392.
[38] Emura T, Nakatochi M, Murotani K, et al. A joint frailty-copula model between tumour progression and death for meta-analysis. Stat Method Med Res. 2017b;26:2649–2666.
[39] Peng M, Xiang L, Wang S. Semiparametric regression analysis of clustered survival data with semi-competing risks. Comput Stat Data Anal. 2018;124:53–70.
[40] Schepsmeier U, Stöber J. Derivatives and Fisher information of bivariate copulas. Stat Pap. 2014;55:525–542.
[41] Shih JH, Emura T. Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula. Comput Stat. 2018;33:1293–1323.