Comput Stat (2018) 33:1293–1323
https://doi.org/10.1007/s00180-018-0804-0

ORIGINAL PAPER

Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula

Jia-Han Shih[1] · **Takeshi Emura**[1]

Received: 10 June 2017 / Accepted: 6 March 2018 / Published online: 16 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract Many existing latent failure time models for competing risks do not provide closed form expressions of sub-distribution functions. This paper suggests a generalized FGM copula models with the Burr III failure time distribution such that the sub-distribution functions have closed form expressions. Under the suggested model, we develop a likelihood-based inference method along with its computational tools and asymptotic theory. Based on the expressions of the sub-distribution functions, we propose goodness-of-fit tests. Simulations are conducted to examine the performance of the proposed methods. A real data from the reliability analysis of the radio transmitter-receivers are analyzed to illustrate the proposed methods. The computational programs are made available in the R package *GFGM.copula*.

Keywords Bivariate survival analysis · Burr III distribution · Copula · Parametric bootstrap · Reliability

1 Introduction

The issue of competing risks arises when researchers deal with multiple failure times. In reliability research, a manufactured product has multiple causes of failure, where only the first-occurring failure time is observable in a series system (Meeker and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00180-018-0804-0) contains supplementary material, which is available to authorized users.

✉ Takeshi Emura
takeshiemura@gmail.com

Jia-Han Shih
tommy355097@gmail.com

[1] Graduate Institute of Statistics, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan City 32001, Taiwan

Escobar 1998; Crowder 2001). In many cases, a goal of competing risks analyses is to make inference on the distributions of the component failure times without assuming their independence.

This paper focuses on parametric approaches to deal with dependent competing risks based on "latent failure time models" (Crowder 2001). In the literature, there are many parametric approaches for latent failure time models such as the bivariate Weibull model (Moeschberger 1974), bivariate normal model (Basu and Ghosh 1978), bivariate log-normal model (Fan and Hsu 2015), and bivariate piecewise exponential model (Staplin et al. 2015).

An alternative to these well-known bivariate models is a copula model (Nelsen 2006; Durante and Sempi 2015; Emura and Chen 2018) in which marginal failure time models are flexibly chosen. Escarela and Carrière (2003) considered the Frank copula with the three-parameter Burr XII margins, where the Burr XII distribution includes the log-logistic, the Pareto II (Lomax), and the Weibull distributions as special cases. Emura et al. (2015) applied the Clayton copula with the margins approximated by a five-parameter spline. Hsu et al. (2016) used the Clayton copula with the margins specified by a class of log-location-scale models. Emura and Michimae (2017) adopted the Clayton or Joe copula with the margins approximated by piecewise exponential models. All these methods utilize likelihood-based parametric inference based on competing risks data.

Aforementioned models, including the popular bivariate models and the Clayton, Frank, and Joe copula models, do not yield closed form expressions for sub-distribution functions. This drawback does not relate to parameter estimation, but it causes inconvenience for purpose of model-diagnostic procedures or goodness-of-fit tests, fundamental tools in parametric analyses. Furthermore, no paper seems to examine the regularity conditions for these models to support the asymptotic theory. In contrast, the asymptotic theory is more rigorously analyzed under some semi-parametric approaches (Rivest and Wells 2001; Braekers and Veraverbeke 2005; de Uña-Álvarez and Veraverbeke 2013, 2017).

In this context, we consider the generalized Farlie–Gumbel–Morgenstern (FGM) copula with the Burr III marginal distribution such that the sub-distribution functions have closed form expressions (Sect. 2). Under the proposed model, we develop a likelihood-based inference method, including its computational algorithms and goodness-of-fit tests (Sect. 3). We provide more convincing asymptotic theory for the proposed method than those for the existing parametric methods (Sect. 4). We provide numerical studies based on Monte-Carlo simulations (Sect. 5) and a real data example based on the reliability analysis of the radio transmitter-receivers (Sect. 6). Some discussions are given in Sect. 7.

2 Competing risks models

2.1 Generalized FGM copula and Burr III distribution

The FGM copula was first introduced by Morgenstern (1956), which can even be traced back to Eyraud (1936). The FGM copula is defined as

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

where $-1 \le \theta \le 1$. The function $C_\theta(u,\ v)$ is a bivariate distribution function where the marginal distributions are the uniform on [0, 1] and the level of dependence is determined by θ. Positive (or negative) dependence corresponds to $\theta > 0$ (or $\theta < 0$), and independence corresponds to $\theta = 0$. Due to its simple form, the FGM copula has been used to illustrate some theoretical results or aging properties of multivariate lifetime models. A recent example includes Navarro (2016) who demonstrated their results on residual lifetimes of coherent systems by the trivariate FGM copula.

Bairamov and Kotz (2002) considered the generalized FGM copula defined as

$$C_\theta(u,\ v) = uv\{1 + \theta(1-u^p)^q(1-v^p)^q\},\quad p \ge 1,\quad q > 1,\quad 0 \le u,\ v \le 1,$$

where the admissible range of θ is

$$-\min\left\{1,\ \frac{1}{p^{2q}}\left(\frac{1+pq}{q-1}\right)^{2q-2}\right\} \le \theta \le \frac{1}{p^q}\left(\frac{1+pq}{q-1}\right)^{q-1}. \tag{1}$$

The restriction $q > 1$ is imposed since only $\theta = 0$ is admissible for $0 < q < 1$ (Bairamov and Kotz 2002). This restriction agrees with Bairamov and Bayramoglu (2013). The generalized FGM copula is symmetric (exchangeable) and preserves the tractability of the FGM copula, while providing closed form expressions of Kendall's tau, Spearman's rho, and other measures that helps us to interpret the meaning of the copula parameters (Amini et al. 2011; Domma and Giordano 2013, 2016; Shih and Emura 2016).

A review of Bairamov and Bayramoglu (2013) reveals that there are several different versions of the generalized FGM copulas, such as those defined by Huang and Kotz (1999) and non-symmetric FGM copulas. However, it would suffice to consider symmetric copulas since non-symmetric copulas are rarely used in practical applications.

Let X and Y be continuous failure times having marginal distribution functions F_1 and F_2. We assume that the joint distribution function $F(x,\ y) = \Pr(X \le x,\ Y \le y)$ is

$$\begin{aligned} F(x,\ y) &= C_\theta\{F_1(x),\ F_2(y)\} \\ &= F_1(x)F_2(y)[1 + \theta\{1 - F_1(x)^p\}^q\{1 - F_2(y)^p\}^q]. \end{aligned} \tag{2}$$

Gumbel (1960) used the exponential distributions for F_1 and F_2 under the FGM copula. Instead, we adopt the Burr III distributions defined as

$$F_{1,(\alpha,\gamma)}(x) = (1 + x^{-\gamma})^{-\alpha},\quad x > 0,\quad F_{2,(\beta,\gamma)}(y) = (1 + y^{-\gamma})^{-\beta},\quad y > 0,$$

where $(\alpha,\ \beta,\ \gamma)$ are all positive shape parameters.

The Burr III distribution belongs to the Burr system (Burr 1942) and is widely used in real applications. It has been applied to financial study (Sherrick et al. 1996), environmental studies (Lindsay et al. 1996; Shao 2000), and reliability theory (Mokhlis 2005). Especially, under the FGM copula with the Burr III margins, Domma and Giordano (2013) derived the explicit formula of a reliability measure which is written as

$$\Pr(Y < X) = \iint_{y<x} \frac{\partial^2}{\partial x \partial y} F(x,\ y) dx dy = \frac{\alpha}{\alpha+\beta} + \theta \frac{\alpha\beta(\alpha-\beta)}{(\alpha+\beta)(2\alpha+\beta)(\alpha+2\beta)}.$$

Shih and Emura (2016) further extended their result to a "truncated reliability measure"

$$\Pr(Y \le t,\ Y < X) = H_\gamma(t)^\beta - \frac{\beta}{\alpha+\beta} H_\gamma(t)^{\alpha+\beta} - \theta\beta \left\{ \frac{1}{\alpha+\beta} H_\gamma(t)^{\alpha+\beta} - \frac{1}{2\alpha+\beta} H_\gamma(t)^{2\alpha+\beta} - \frac{2}{\alpha+2\beta} H_\gamma(t)^{\alpha+2\beta} + \frac{1}{\alpha+\beta} H_\gamma(t)^{2\alpha+2\beta} \right\},$$

where $H_\gamma(t) = (1 + t^{-\gamma})^{-1}$. The results can also be extended to the generalized FGM copula. These properties make the Burr III distribution as an attractive model for reliability, competing risks, and truncated data analyses.

2.2 Bivariate competing risks models

In bivariate competing risks analysis, observed variables are the failure time T and the cause indicator J taking values $J = 1$ (cause 1 failure) or $J = 2$ (cause 2 failure). Define

$$F(1,\ t) = \Pr(J = 1,\ T \le t) \quad \text{and} \quad F(2,\ t) = \Pr(J = 2,\ T \le t).$$

They are the *sub-distribution functions* for $F_T(t) = \Pr(T \le t)$ such that $F(1,\ t) + F(2,\ t) = F_T(t)$. In competing risks analysis, the sub-distribution functions play a fundamental role as it is empirically estimable, and easy to interpret (Crowder 2001; Escarela and Carrière 2003; Lawless 2003; Bakoyannis and Touloumi 2012).

The latent failure time model for bivariate competing risks considers two latent failure times X and Y corresponding to cause 1 and cause 2 failures, respectively (Chap 3 of Crowder 2001). One can observe the first occurring event time $T = \min(X,\ Y)$ and the cause indicator J taking values $J = 1$ (if $X \le Y$) or $J = 2$ (if $Y < X$). We assume that X and Y are continuous and non-negative random variables following the generalized FGM copula model in Eq. (2).

For instance, consider a life test of a series system having two components, say components 1 and 2. The system fails if either component 1 or 2 fails. At the time of failure, one can observe the first occurring component failure time and its failure cause. If the system does not fail at the end of the life test, one can only know that the two component failure times are greater than the time. See Sect. 6 for more detailed illustration on a similar example.

Under the latent failure time model, the sub-distribution functions can be rewritten as

$$F(j,\ t) = \Pr(J = j,\ T \le t) = \int_0^t f(j,\ z)dz, \quad j = 1,\ 2.$$

where $f(1,\ t) = -\partial \bar{F}(x,\ y)/\partial x|_{x=y=t}$ and $f(2,\ t) = -\partial \bar{F}(x,\ y)/\partial y|_{x=y=t}$ are known as the *sub-density functions*, and

$$\bar{F}(x,\ y) = 1 - F_1(x) - F_2(y) + F_1(x)F_2(y)[1 + \theta\{1 - F_1(x)^p\}^q\{1 - F_2(y)^p\}^q]$$

is the joint survival function.

In the existing copula-based latent failure time methods for competing risks data, the sub-distribution functions have not been the main tool for estimation. Instead, existing likelihood-based estimation methods utilize the cause-specific hazard function in their likelihood functions (e.g., Escarela and Carrière 2003; Chen 2010; Emura and Chen 2016; Emura and Chen 2018). However, the sub-distribution functions are potentially useful for model-diagnostic procedures (Escarela and Carrière 2003) or goodness-of-fit tests.

Escarela and Carrière (2003) proposed the survival copula model

$$\Pr(X > x,\ Y > y) = \bar{F}(x,\ y) = C_\theta\{1 - F_1(x),\ 1 - F_2(y)\},$$

where

$$C_\theta(u,\ v) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right\}, \quad \theta \ne 0$$

is the Frank copula, $F_j(t) = 1 - \exp\{-(\lambda_j t\ )^{\alpha_j}\}$, $\lambda_j,\ \alpha_j > 0$, $j = 1,\ 2$ are the Weibull distribution functions. To assess the goodness-of-fit of their proposed model, they compared the empirical sub-distribution functions with the model-based sub-distribution functions:

$$F(j,\ t) = \int_0^t \alpha_j\lambda_j(\lambda_j z\ )^{\alpha_j - 1}\{1 - F_j(z)\}\frac{\exp[-\theta\{1 - F_j(z)\}]\{\exp[-\theta\{1 - F_{3-j}(z)\}] - 1\}}{(e^{-\theta} - 1)\exp\{-\theta\bar{F}(z,\ z)\}}dz$$

for $j = 1, 2$. Escarela and Carrière (2003) pointed out that the above sub-distribution does not have a closed form. They obtained it numerically by the S-PLUS *integrate* function. Many common bivariate distributions do not yield closed form sub-distribution functions.

Under the generalized FGM copula model, Shih and Emura (2016) derived the explicit expressions of sub-densities and sub-distribution functions under the Burr III margins, defined as $F_{1,(\alpha,\gamma)}(x) = (1 + x^{-\gamma})^{-\alpha}$ and $F_{2,(\beta,\gamma)}(y) = (1 + y^{-\gamma})^{-\beta}$, where $(\alpha,\ \beta,\ \gamma)$ are positive shape parameters. The sub-density functions can be expressed as

$$f_{\boldsymbol{\varphi}}(1,\ t) = \alpha\gamma\, t^{-\gamma-1} H_\gamma(t)^{\alpha+1} - \alpha\gamma\, t^{-\gamma-1} H_\gamma(t)^{\alpha+\beta+1} - \theta\alpha\gamma \sum_{i=0}^{q}\sum_{j=0}^{q-1} \binom{q}{i}\binom{q-1}{j} (-1)^{i+j} H_\gamma(t)^{\alpha(pj+1)+\beta(pi+1)+1}\{1-(1+pq)H_\gamma(t)^{\alpha p}\} t^{-\gamma-1} \tag{3}$$

and

$$f_{\boldsymbol{\varphi}}(2,\ t) = \beta\gamma\, t^{-\gamma-1} H_\gamma(t)^{\beta+1} - \beta\gamma\, t^{-\gamma-1} H_\gamma(t)^{\alpha+\beta+1} - \theta\beta\gamma \sum_{i=0}^{q}\sum_{j=0}^{q-1} \binom{q}{i}\binom{q-1}{j} (-1)^{i+j} H_\gamma(t)^{\alpha(pi+1)+\beta(pj+1)+1}\{1-(1+pq)H_\gamma(t)^{\beta p}\} t^{-\gamma-1}, \tag{4}$$

where $\boldsymbol{\varphi} = (\alpha,\ \beta,\ \gamma)$, $H_\gamma(t) = (1+t^{-\gamma})^{-1}$, $p \geq 1$, $q \in \{1,\ 2,\ \ldots\}$, and θ in the range of Eq. (1). The sub-distribution functions can be expressed as

$$F_{\boldsymbol{\varphi}}(1,\ t) = H_\gamma(t)^{\alpha} - \frac{\alpha}{\alpha+\beta} H_\gamma(t)^{\alpha+\beta} - \theta\alpha \sum_{i=0}^{q}\sum_{j=0}^{q-1} \binom{q}{i}\binom{q-1}{j} (-1)^{i+j} \left\{ \frac{H_\gamma(t)^{\alpha(pj+1)+\beta(pi+1)}}{\alpha(pj+1)+\beta(pi+1)} - \frac{(1+pq)H_\gamma(t)^{\alpha(pj+p+1)+\beta(pi+1)}}{\alpha(pj+p+1)+\beta(pi+1)} \right\}$$

and

$$F_{\boldsymbol{\varphi}}(2,\ t) = H_\gamma(t)^{\beta} - \frac{\beta}{\alpha+\beta} H_\gamma(t)^{\alpha+\beta} - \theta\beta \sum_{i=0}^{q}\sum_{j=0}^{q-1} \binom{q}{i}\binom{q-1}{j} (-1)^{i+j} \left\{ \frac{H_\gamma(t)^{\alpha(pi+1)+\beta(pj+1)}}{\alpha(pi+1)+\beta(pj+1)} - \frac{(1+pq)H_\gamma(t)^{\alpha(pi+1)+\beta(pj+p+1)}}{\alpha(pi+1)+\beta(pj+p+1)} \right\}.$$

However, Shih and Emura (2016) did not consider statistical inference under the model.

3 Proposed method

This section considers parameter estimation and proposes goodness-of-fit tests under the generalized FGM copula model with the Burr III margins.

3.1 Maximum likelihood estimation

The generalized FGM copula model with the Burr III margins is defined as

$$F_{\boldsymbol{\varphi}}(x,\ y) = (1+x^{-\gamma})^{-\alpha}(1+y^{-\gamma})^{-\beta}[1+\theta\ \{1-(1+x^{-\gamma})^{-\alpha p}\}^{q}\{1-(1+y^{-\gamma})^{-\beta p}\}^{q}], \tag{5}$$

where $\boldsymbol{\varphi} = (\alpha,\ \beta,\ \gamma)$. The marginal distribution functions for X and Y are

$$F_{1,(\alpha,\gamma)}(x) = F_{\boldsymbol{\varphi}}(x,\ \infty) = (1+x^{-\gamma})^{-\alpha} \quad \text{and} \quad F_{2,(\beta,\gamma)}(y) = F_{\boldsymbol{\varphi}}(\infty,\ y) = (1+y^{-\gamma})^{-\beta}.$$

The joint survival function is

$$\bar{F}_{\boldsymbol{\varphi}}(x,\ y) = 1 - F_{1,(\alpha,\gamma)}(x) - F_{2,(\beta,\gamma)}(y) + F_{\boldsymbol{\varphi}}(x,\ y). \tag{6}$$

Table 1 Three different observation patterns under bivariate competing risks

Observation patterns	T_i	δ_i	δ_i^*	Likelihood contribution
Cause 1 failure	X_i	1	0	$\Pr(X_i = T_i,\ Y_i > T_i) = f_{\boldsymbol{\varphi}}(1,\ T_i)$
Cause 2 failure	Y_i	0	1	$\Pr(X_i > T_i,\ Y_i = T_i) = f_{\boldsymbol{\varphi}}(2,\ T_i)$
Censoring	C_i	0	0	$\Pr(X_i > T_i,\ Y_i > T_i) = \bar{F}_{\boldsymbol{\varphi}}(T_i,\ T_i)$

In the presence of competing risks, one can observe only one of X and Y for each subject. This implies that one cannot use correlation statistics such as Pearson's correlation and Kendall's tau to estimate copula parameters $(p,\ q,\ \theta)$. Indeed, without any model assumption on X and Y, one cannot identify their correlation structure (Tsiatis 1975).

If some parametric or semi-parametric models are imposed on X and Y, the model becomes identifiable. Such models include one- or two-parameter lifetime models (David and Moeschberger 1978; Basu and Ghosh 1978), parametric Weibull regression models (Escarela and Carrière 2003), the semi-parametric Cox and AFT models (Heckman and Honore 1989), and copula models (Zheng and Klein 1995). However, even if the model is identifiable, the estimation of the dependence parameter remains a difficult task since the competing risks data provide little information about dependence between X and Y.

To identify the copula parameters $(p,\ q,\ \theta)$ under the model (5), we will follow Escarela and Carrière (2003) who adopt a profile-likelihood approach. The implementation of the profile likelihood approach under the model (5) will be described in Sect. 6. The rest of this section develops an estimation method for the marginal parameter $\boldsymbol{\varphi} = (\alpha,\ \beta,\ \gamma)$ given the copula parameters $(p,\ q,\ \theta)$.

Let $(X_i,\ Y_i,\ C_i)$, $i = 1,\ 2,\ \ldots,\ n$, be i.i.d. triplets, where $(X_i,\ Y_i)$ follows Eq. (5) and C_i is the independent censoring time. Type I censoring corresponds to the case where $C_i = w$ for all i, where w is the duration of the life test. Denote $T_i = \min(X_i,\ Y_i,\ C_i)$ as the observed failure time, $\delta_i = \mathbf{I}(T_i = X_i)$ as the indicator of failure cause 1, and $\delta_i^* = \mathbf{I}(T_i = Y_i)$ as the indicator of failure cause 2. The data consist of $(T_i,\ \delta_i,\ \delta_i^*)$ for $i = 1,\ 2,\ \ldots,\ n$ (Table 1).

As in Lawless (2003, p. 435), the log-likelihood function is

$$\ell_n(\boldsymbol{\varphi}) = \sum_{i=1}^{n} \delta_i \log f_{\boldsymbol{\varphi}}(1,\ T_i) + \sum_{i=1}^{n} \delta_i^* \log f_{\boldsymbol{\varphi}}(2,\ T_i) + \sum_{i=1}^{n} (1 - \delta_i - \delta_i^*) \log \bar{F}_{\boldsymbol{\varphi}}(T_i,\ T_i),$$

where $f_{\boldsymbol{\varphi}}(1,\ t)$, $f_{\boldsymbol{\varphi}}(2, t)$, and $\bar{F}_{\boldsymbol{\varphi}}(t, t)$ are explicitly written in Eqs. (3), (4), and (6), respectively. The maximum likelihood estimator (MLE) is

$$\hat{\boldsymbol{\varphi}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \arg\max_{\boldsymbol{\varphi} \in \Theta} \ell_n(\boldsymbol{\varphi}), \quad \Theta = (0, \infty) \times (0, \infty) \times (0, \infty).$$

To obtain the MLE, we use the derivatives of $\ell_n(\boldsymbol{\varphi})$ with respect to $\boldsymbol{\varphi} = (\alpha, \beta, \gamma)$.

Let φ_j be the j-th element of $\boldsymbol{\varphi} = (\alpha, \beta, \gamma)$, $j = 1, 2, 3$. The first-order derivatives of the log-likelihood function (score functions) are

$$\frac{\partial \ell_n(\boldsymbol{\varphi})}{\partial \phi_j} = \sum_{i=1}^{n} \left\{ \delta_i \frac{1}{f_{\boldsymbol{\varphi}}(1, T_i)} \frac{\partial f_{\boldsymbol{\varphi}}(1, T_i)}{\partial \phi_j} \right\} + \sum_{i=1}^{n} \left\{ \delta_i^* \frac{1}{f_{\boldsymbol{\varphi}}(2, T_i)} \frac{\partial f_{\boldsymbol{\varphi}}(2, T_i)}{\partial \phi_j} \right\} + \sum_{i=1}^{n} \left\{ (1 - \delta_i - \delta_i^*) \frac{1}{\bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)} \frac{\partial \bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)}{\partial \phi_j} \right\}, \quad j = 1, 2, 3,$$

where the formulas for $\partial f_{\boldsymbol{\varphi}}(1, t)/\partial \phi_j$, $\partial f_{\boldsymbol{\varphi}}(2, t)/\partial \phi_j$, and $\partial \bar{F}_{\boldsymbol{\varphi}}(t, t)/\partial \phi_j$, $j = 1, 2, 3$ are given in Supplementary Material.

The second-order derivatives of the log-likelihood function (Hessians) are

$$\begin{aligned}\frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \phi_j \partial \phi_k} &= \sum_{i=1}^{n} \left[\delta_i \left\{ \frac{1}{f_{\boldsymbol{\varphi}}(1, T_i)} \frac{\partial^2 f_{\boldsymbol{\varphi}}(1, T_i)}{\partial \phi_j \partial \phi_k} - \frac{1}{f(1, T_i)^2} \frac{\partial f_{\boldsymbol{\varphi}}(1, T_i)}{\partial \phi_j} \frac{\partial f_{\boldsymbol{\varphi}}(1, T_i)}{\partial \phi_k} \right\} \right] \\ &+ \sum_{i=1}^{n} \left[\delta_i^* \left\{ \frac{1}{f_{\boldsymbol{\varphi}}(2, T_i)} \frac{\partial^2 f_{\boldsymbol{\varphi}}(1, T_i)}{\partial \phi_j \partial \phi_k} - \frac{1}{f_{\boldsymbol{\varphi}}(2, T_i)^2} \frac{\partial f_{\boldsymbol{\varphi}}(2, T_i)}{\partial \phi_j} \frac{\partial f_{\boldsymbol{\varphi}}(2, T_i)}{\partial \phi_k} \right\} \right] \\ &+ \sum_{i=1}^{n} \left[(1 - \delta_i - \delta_i^*) \left\{ \frac{1}{\bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)} \frac{\partial^2 \bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)}{\partial \phi_j \partial \phi_k} - \frac{1}{\bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)^2} \frac{\partial \bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)}{\partial \phi_j} \frac{\partial \bar{F}_{\boldsymbol{\varphi}}(T_i, T_i)}{\partial \phi_k} \right\} \right], \\ & j, \ k = 1, \ 2, \ 3,\end{aligned}$$

where the formulas for $\partial^2 f_{\boldsymbol{\varphi}}(1, t)/\partial \phi_j \partial \phi_k$, $\partial^2 f_{\boldsymbol{\varphi}}(2, t)/\partial \phi_j \partial \phi_k$, and $\partial^2 \bar{F}_{\boldsymbol{\varphi}}(t, t)/\partial \phi_j \partial \phi_k$, $j, k = 1, 2, 3$ are given in Supplementary Material.

With the score functions and Hessians, one can perform the Newton–Raphson (NR) algorithm. In our experience, the NR algorithm ascertains the MLE within 5 or 6 iterations. On the other hand, the NR algorithm is sensitive to the initial value (Knight 2000). Our close inspection revealed that the NR algorithm gives erroneous results when $f_{\boldsymbol{\varphi}}(1, T_i) \approx 0$, $f_{\boldsymbol{\varphi}}(2, \ T_i) \approx 0$, or $\bar{F}_{\boldsymbol{\varphi}}(T_i, \ T_i) \approx 0$ for some i during iterations. These irregular cases imply that the NR algorithm does not converge due to a wrong initial value. To solve these problems, it suffices to apply the randomized NR algorithm (Hu and Emura 2015).

Algorithm 1: Randomized Newton–Raphson algorithm
Let D, c, ε_0, ε_1, ε_2, r_1, r_2, and r_3 be some positive tuning parameters.

Step1 Set an initial value $(\alpha^{(0)}, \ \beta^{(0)}, \ \gamma^{(0)}) = (\bar{X}, \ \bar{Y}, \ c)$, where

$$\bar{X} = \sum_{i=1}^{n} \delta_i T_i / \sum_{i=1}^{n} \delta_i \quad \text{and} \quad \bar{Y} = \sum_{i=1}^{n} \delta_i^* T_i / \sum_{i=1}^{n} \delta_i^*.$$

Step2 Calculate $Err^{(k+1)} = \max\{|\alpha^{(k+1)} - \alpha^{(k)}|,\ |\beta^{(k+1)} - \beta^{(k)}|,\ |\gamma^{(k+1)} - \gamma^{(k)}|\}$, where

$$\begin{bmatrix} \alpha^{(k+1)} \\ \beta^{(k+1)} \\ \gamma^{(k+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(k)} \\ \beta^{(k)} \\ \gamma^{(k)} \end{bmatrix} - \left. \begin{bmatrix} \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \alpha^2} & \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \alpha \partial \gamma} \\ \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \beta^2} & \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \beta \partial \gamma} \\ \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \alpha \partial \gamma} & \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \beta \partial \gamma} & \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \gamma^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \ell_n(\boldsymbol{\varphi})}{\partial \alpha} \\ \frac{\partial \ell_n(\boldsymbol{\varphi})}{\partial \beta} \\ \frac{\partial \ell_n(\boldsymbol{\varphi})}{\partial \gamma} \end{bmatrix} \right|_{\alpha=\alpha^{(k)},\ \beta=\beta^{(k)},\ \gamma=\gamma^{(k)}}$$

- If $Err^{(k+1)} < \varepsilon_0$, stop the algorithm and the MLE is $\hat{\boldsymbol{\varphi}} = (\alpha^{(k+1)},\ \beta^{(k+1)},\ \gamma^{(k+1)})$.
- If $Err^{(k+1)} > D$, $\min\{\alpha^{(k+1)},\ \beta^{(k+1)},\ \gamma^{(k+1)}\} < \varepsilon_1$, $f_{\boldsymbol{\varphi}}(1,\ T_i) < \varepsilon_2$, $f_{\boldsymbol{\varphi}}(2,\ T_i) < \varepsilon_2$, or $\bar{F}_{\boldsymbol{\varphi}}(T_i,\ T_i) < \varepsilon_2$ for some i, stop the algorithm and return to Step 1 with the initial value being replaced by $(\alpha^{(0)}e^{u_1},\ \beta^{(0)}e^{u_2},\ \gamma^{(0)}e^{u_3})$, where $u_i \sim U(-r_i,\ r_i,$ $i = 1,\ 2,\ 3$.

Remark 1 The choice $\alpha^{(0)} = \bar{X}$ comes from the formula $E(X) = \alpha B(\alpha + 1/\gamma,\ 1 - 1/\gamma)$. The scale of α may be reflected by the value of $\bar{X}$. Similarly, we set $\beta^{(0)} = \bar{Y}$.

Remark 2 We suggest trying several values for c. For illustration, we run Algorithm 1 on one simulated dataset of $n = 100$ under $c = 1,\ 2,\ 3,\ 4$. Table 2 shows that the choice of c only influences the number of iterations and does not influence parameter estimates.

Remark 3 We set $c = 4$ and $D = 50$ for the simulation (Sect. 5) while $c = 1$ and $D = 5000$ for the data analysis (Sect. 6). The other tuning parameters are set to be $r_1 = r_2 = r_3 = 1$, $\varepsilon_0 = 10^{-5}$, $\varepsilon_1 = 10^{-10}$, and $\varepsilon_2 = 10^{-300}$ for both simulation and data analysis.

3.2 Goodness-of-fit tests

We develop goodness-of-fit tests for examining the fit of the model (5). Our tests are inspired by a graphical tool proposed by Escarela and Carrière (2003). The tests consist of three estimators of the sub-distribution functions: (i) parametric estimator, (ii) semi-parametric estimator, and (iii) non-parametric estimator. If the three estimators are

Table 2 The results of the randomized NR algorithm (Algorithm 1) under different initial values

Initial values	$\hat{\alpha}$ $(\alpha = 3)$	$\hat{\beta}$ $(\beta = 2)$	$\hat{\gamma}$ $(\gamma = 7)$	The number of iterations	The number of randomizations
$c = 1$	3.400	1.995	6.899	7	64
$c = 2$	3.400	1.995	6.899	6	8
$c = 3$	3.400	1.995	6.899	6	1
$c = 4$	3.400	1.995	6.899	7	0

The results are based on one simulated dataset with $n = 100$

close to one another, we conclude that there is no evidence against the model. Escarela and Carrière (2003) considered only the graphical comparison between (i) and (iii) without significance tests.

(i) Parametric estimation of sub-distribution functions

Under the generalized FGM copula model with Burr III margins in Eq. (5), the sub-distribution functions have closed forms (Sect. 2.2). Hence, if the model (5) is correct, the consistent estimators of the sub-distribution functions are

$$F_{\hat{\boldsymbol{\varphi}}}(1,\ t) = H_{\hat{\gamma}}(t)^{\hat{\alpha}} - \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} H_{\hat{\gamma}}(t)^{\hat{\alpha}+\hat{\beta}} - \theta\hat{\alpha}\sum_{i=0}^{q}\sum_{j=0}^{q-1}\binom{q}{i}\binom{q-1}{j}(-1)^{i+j}\left\{\frac{H_{\hat{\gamma}}(t)^{\hat{\alpha}(pj+1)+\hat{\beta}(pi+1)}}{\hat{\alpha}(pj+1)+\hat{\beta}(pi+1)} - \frac{(1+pq)H_{\hat{\gamma}}(t)^{\hat{\alpha}(pj+p+1)+\hat{\beta}(pi+1)}}{\hat{\alpha}(pj+p+1)+\hat{\beta}(pi+1)}\right\}$$

and

$$F_{\hat{\boldsymbol{\varphi}}}(2,\ t) = H_{\hat{\gamma}}(t)^{\hat{\beta}} - \frac{\hat{\beta}}{\hat{\alpha}+\hat{\beta}} H_{\hat{\gamma}}(t)^{\hat{\alpha}+\hat{\beta}} - \theta\hat{\beta}\sum_{i=0}^{q}\sum_{j=0}^{q-1}\binom{q}{i}\binom{q-1}{j}(-1)^{i+j}\left\{\frac{H_{\hat{\gamma}}(t)^{\hat{\alpha}(pi+1)+\hat{\beta}(pj+1)}}{\hat{\alpha}(pi+1)+\hat{\beta}(pj+1)} - \frac{(1+pq)H_{\hat{\gamma}}(t)^{\hat{\alpha}(pi+1)+\hat{\beta}(pj+p+1)}}{\hat{\alpha}(pi+1)+\hat{\beta}(pj+p+1)}\right\},$$

where $H_{\hat{\gamma}}(t) = (1+t^{-\hat{\gamma}})^{-1}$ and $\hat{\boldsymbol{\varphi}} = (\hat{\alpha},\ \hat{\beta},\ \hat{\gamma})$ is the MLE.

(ii) Semi-parametric estimation of sub-distribution functions

For semi-parametric estimation, we consider the generalized FGM copula model with margins being unspecified. Here, we approximate the hazard function of X as

$$h_{1,\mathbf{g}_1}(x) = \sum_{\ell=1}^{L_X} g_{1\ell} M_\ell(x),$$

where $M_\ell(x)$, $\ell = 1,\ 2,\ \ldots,\ L_X$, are the cubic M-spline bases (Ramsay 1988) and $\mathbf{g}_1 = (g_{11},\ \ldots,\ g_{1L_X})$ are unknown parameters with $g_{1\ell} \ge 0$ for $\ell = 1,\ 2,\ \ldots,\ L_X$. The cumulative hazard function of X is then approximated by

$$H_{1,\mathbf{g}_1}(x) = \sum_{\ell=1}^{L_X} g_{1\ell} I_\ell(x),$$

where $I_\ell(x)$ is the integration of $M_\ell(x)$, called as the I-spline bases (Ramsay 1988). The hazard and cumulative hazard functions of Y are similarly approximated as

$$h_{2,\mathbf{g}_2}(y) = \sum_{\ell=1}^{L_Y} g_{2\ell} M_\ell(y) \quad \text{and} \quad H_{2,\mathbf{g}_2}(y) = \sum_{\ell=1}^{L_Y} g_{2\ell} I_\ell(y),$$

where $\mathbf{g}_2 = (g_{21}, \ldots, g_{2L_Y})$ with $g_{2\ell} \geq 0$ for $\ell = 1, 2, \ldots, L_Y$. With the above definitions, the approximate marginal distribution and density functions of X and Y can be written as

$$F_{j,\mathbf{g}_j}(x) = 1 - \exp\{-H_{j,\mathbf{g}_j}(x)\}, \ f_{j,\mathbf{g}_j}(x) = h_{j,\mathbf{g}_j}(x)\exp\{-H_{j,\mathbf{g}_j}(x)\}, \ j = 1, 2.$$

Letting $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2)$, the semi-parametric model is written as

$$F_{\mathbf{g}}(x, y) = \Pr(X \leq x, Y \leq y) = F_{1,\mathbf{g}_1}(x)F_{2,\mathbf{g}_2}(y)[1 + \theta \{1 - F_{1,\mathbf{g}_1}(x)^p\}^q\{1 - F_{2,\mathbf{g}_2}(y)^p\}^q],$$

Let $\hat{\mathbf{g}}$ be the MLE under this model. We use the M- and I-spline bases of $L_X = L_Y = 5$ derived in Emura et al. (2015) and Emura and Chen (2018) which are available in the R *joint.Cox* package (Emura 2017).

By Theorem 6 in Shih and Emura (2016), we estimate the sub-distribution functions as

$$\begin{aligned} F_{\hat{\mathbf{g}}}(1, t) &= F_{1,\hat{\mathbf{g}}_1}(t) - \int_0^t F_{2,\hat{\mathbf{g}}_2}(z)f_{1,\hat{\mathbf{g}}_1}(z)dz \\ &\quad - \theta\sum_{i=0}^{q}\sum_{j=0}^{q-1}\binom{q}{i}\binom{q-1}{j}(-1)^{i+j}\int_0^t F_{2,\hat{\mathbf{g}}_2}(z)^{pi+1}F_{1,\hat{\mathbf{g}}_1}(z)^{pj}\{1-(1+pq)F_{1,\hat{\mathbf{g}}_1}(z)^p\}f_{1,\hat{\mathbf{g}}_1}(z)dz \end{aligned}$$

and

$$\begin{aligned} F_{\hat{\mathbf{g}}}(2, t) &= F_{2,\hat{\mathbf{g}}_2}(t) - \int_0^t F_{1,\hat{\mathbf{g}}_1}(z)f_{2,\hat{\mathbf{g}}_2}(z)dz \\ &\quad - \theta\sum_{i=0}^{q}\sum_{j=0}^{q-1}\binom{q}{i}\binom{q-1}{j}(-1)^{i+j}\int_0^t F_{1,\hat{\mathbf{g}}_1}(z)^{pi+1}F_{2,\hat{\mathbf{g}}_2}(z)^{pj}\{1-(1+pq)F_{2,\hat{\mathbf{g}}_2}(z)^p\}f_{2,\hat{\mathbf{g}}_2}(z)dz, \end{aligned}$$

where the integrals are computed by the R *integrate* function.

(iii) Non-parametric estimation of sub-distribution functions

For non-parametric estimation, we let $T_{(1)} < T_{(2)} < \cdots < T_{(k)}$ be distinct uncensored times. Then, the non-parametric estimators of sub-distribution functions (Lawless 2003, p.437-439) are

$$\hat{F}(1, t) = \sum_{i:\, T_{(i)} \leq t}\hat{S}(t)\frac{d_{i1}}{n_i} \quad \text{and} \quad \hat{F}(2, t) = \sum_{i:\, T_{(i)} \leq t}\hat{S}(t)\frac{d_{i2}}{n_i},$$

where

$$\hat{S}(t) = \prod_{i: T_{(i)} < t}\frac{n_i - d_i}{n_i}, \ n_i = \sum_{j=1}^{n}\mathbf{I}(T_j \geq T_{(i)}), \ d_i = \sum_{j=1}^{n}\mathbf{I}(T_j = T_{(i)}),$$

$$d_{i1} = \sum_{j=1}^{n}\delta_i\mathbf{I}(T_j = T_{(i)}), \ \text{and} \ d_{i2} = \sum_{j=1}^{n}\delta_i^*\mathbf{I}(T_j = T_{(i)}).$$

Based on these estimators for sub-distribution functions, we consider two test hypotheses: (I) The overall model fit and (II) The fit of the generalized FGM copula.

(I) Test for the overall model fit

The null hypothesis for the overall fit is defined as

$$H_0^1 : \Pr(X \le x,\ Y \le y) = C_\theta\{F_{1,(\alpha,\gamma)}(x),\ F_{2,(\beta,\gamma)}(y)\},\quad \text{for some}\quad (\alpha,\ \beta,\ \gamma),$$

where $C_\theta(u,\ v) = uv\{1 + \theta(1 - u^p)^q(1 - v^p)^q\}$, $F_{2,(\alpha,\beta)}(y) = (1 + y^{-\gamma})^{-\beta}$, and $F_{1,(\alpha,\gamma)}(x) = (1 + x^{-\gamma})^{-\alpha}$. The rejection of this hypothesis implies that either the generalized FGM copula or the marginal Burr III model is incorrect. The test statistic for this hypothesis is based on the distance between the parametric estimators and the non-parametric estimators. Specially, we consider the Cramér-von Mises type statistic

$$S_1 = \sum_{i=1}^{n} \delta_i\{F_{\hat{\boldsymbol{\varphi}}}(1,\ T_i) - \hat{F}(1,\ T_i)\}^2 + \sum_{i=1}^{n} \delta_i^*\{F_{\hat{\boldsymbol{\varphi}}}(2,\ T_i) - \hat{F}(2,\ T_i)\}^2.$$

A large value of S_1 rejects H_0^1 (rejects the overall model).

(II) Test for the fit of the generalized FGM copula

The null hypothesis for the generalized FGM copula is

$$H_0^2 : \Pr(X \le x,\ Y \le y) = C_\theta\{F_1(x),\ F_2(y)\},\quad \text{for some}\quad F_1 \quad \text{and}\quad F_2,$$

where $C_\theta(u,\ v) = uv\{1+\theta(1-u^p)^q(1-v^p)^q\}$. The rejection of this hypothesis implies that the generalized FGM copula is incorrect. The test statistic for this hypothesis is based on the distance between the semi-parametric estimators and the non-parametric estimators, namely

$$S_2 = \sum_{i=1}^{n} \delta_i\{F_{\hat{\mathbf{g}}}(1,\ T_i) - \hat{F}(1,\ T_i)\}^2 + \sum_{i=1}^{n} \delta_i^*\{F_{\hat{\mathbf{g}}}(2,\ T_i) - \hat{F}(2,\ T_i)\}^2.$$

A large value of S_2 rejects H_0^2 (rejects the generalized FGM copula).

To obtain level α tests, we apply the parametric bootstrap (Efron and Tibshirani 1993).

Algorithm 2: The goodness-of-fit tests with parametric bootstrap
Let B be a large integer.

Step 1 Generate bootstrap samples $(T_i^{(b)},\ \delta_i^{(b)},\ \delta_i^{*(b)})$, $i = 1,\ 2,\ \ldots,\ n$, $b = 1,\ 2,\ \ldots,\ B$ under the model (5) with $\hat{\boldsymbol{\varphi}} = (\hat{\alpha},\ \hat{\beta},\ \hat{\gamma})$.

Step 2 Based on the bootstrap samples, compute the Cramér–von Mises type statistics

$$S_1^{(b)} = \sum_{i=1}^{n} \delta_i^{(b)} \{F_{\hat{\boldsymbol{\varphi}}^{(b)}}(1,\ T_i^{(b)}) - \hat{F}^{(b)}(1,\ T_i^{(b)})\}^2 + \sum_{i=1}^{n} \delta_i^{*(b)} \{F_{\hat{\boldsymbol{\varphi}}^{(b)}}(2,\ T_i^{(b)}) - \hat{F}^{(b)}(2,\ T_i^{(b)})\}^2$$

and

$$S_2^{(b)} = \sum_{i=1}^{n} \delta_i^{(b)} \{F_{\hat{\mathbf{g}}^{(b)}}(1,\ T_i^{(b)}) - \hat{F}^{(b)}(1,\ T_i^{(b)})\}^2 + \sum_{i=1}^{n} \delta_i^{*(b)} \{F_{\hat{\mathbf{g}}^{(b)}}(2,\ T_i^{(b)}) - \hat{F}^{(b)}(2,\ T_i^{(b)})\}^2,$$

for each $b = 1,\ 2,\ \ldots,\ B$.

Step 3 If the Cramér-von Mises type statistic S_k is greater than the $100 \times (1 - \alpha)$ percent point of $\{S_k^{(b)},\ b = 1,\ 2,\ \ldots,\ B\}$, we reject the null hypothesis H_0^k for $k = 1,\ 2$.

Remark 4 Step 1 requires an algorithm to generate the censoring time $C_i^{(b)}$. In the simulation, we generate $C_i^{(b)} \sim U(0,\ w)$, where $w = \max_{\delta_i = \delta_i^* = 0}(T_i)$. In the real data analysis, we set $C_i^{(b)} = 630$ according to the Type I censoring condition (Sect. 6).

3.3 Software

The computation programs are implemented in our R package, *GFGM.copula* (Shih 2018). R users can calculate the MLE and estimates of the mean failure times with their SEs and CIs. Users can also calculate the sub-distribution functions and the goodness-of-fit statistics.

4 Asymptotic inference

This section develops the asymptotic theory of the MLE and gives the standard error and confidence interval. The regularity conditions are also examined.

4.1 Asymptotic theory

Let $G(t) = \Pr(C > t)$ and $g(t) = -dG(t)/dt$ be the survival function and density function of the independent censoring time C. Then the joint density of $(T,\ \Delta,\ \Delta^*)$, where $T = \min(X,\ Y,\ C)$, $\Delta = \mathrm{I}(T = X)$, and $\Delta^* = \mathrm{I}(T = Y)$, is defined as

$$f_{\boldsymbol{\varphi}}(t,\ \delta,\ \delta^*) = \{f_{\boldsymbol{\varphi}}(1,\ t)G(t)\}^{\delta}\{f_{\boldsymbol{\varphi}}(2,\ t)G(t)\}^{\delta^*}\{g(t)\bar{F}_{\boldsymbol{\varphi}}(t,\ t)\}^{1-\delta-\delta^*},\quad (t,\ \delta,\ \delta^*) \in \chi,$$

where $\chi = (0,\ \infty) \times \{(1,\ 0),\ (0,\ 1),\ (0,\ 0)\}$ and $\boldsymbol{\varphi} = (\alpha,\ \beta,\ \gamma)$. We assume that G does not depend on $\boldsymbol{\varphi}$. The expectation of any function $h : \mathrm{R}^3 \mapsto \mathrm{R}$ is defined as

$$E_{\boldsymbol{\varphi}}\{h(T,\ \Delta,\ \Delta^*)\} = \int_0^\infty \sum_{(\delta,\delta^*)\in Q} \{h(t,\ \delta,\ \delta^*)f_{\boldsymbol{\varphi}}(t,\ \delta,\ \delta^*)\}dt,$$
$$Q = \{(1,\ 0),\ (0,\ 1),\ (0,\ 0)\}.$$

The Fisher information matrix is denoted as $I(\boldsymbol{\varphi})$ whose elements are

$$I_{jk}(\boldsymbol{\varphi}) = E_{\boldsymbol{\varphi}}\left(\frac{\partial \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial\phi_j}\cdot\frac{\partial \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial\phi_k}\right),\quad j,\ k = 1,\ 2,\ 3.$$

As in Lehmann and Casella (1998, p. 463), we consider the following assumptions:

Assumption (A) There exists an open subset $\omega \subset \Theta$ containing the true parameter, that is $\boldsymbol{\varphi}^0 = (\alpha^0,\ \beta^0,\ \gamma^0) \in \omega$.

Assumption (B) For all $\boldsymbol{\varphi} \in \Theta$, the following equations hold

$$E_{\boldsymbol{\varphi}}\left(\frac{\partial \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial\phi_j}\right) = 0,\quad j = 1,\ 2,\ 3,$$

$$E_{\boldsymbol{\varphi}}\left(\frac{\partial \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial\phi_j}\cdot\frac{\partial \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial\phi_k}\right) = E_{\boldsymbol{\varphi}}\left(-\frac{\partial^2 \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial\phi_j\partial\phi_k}\right),\quad j,\ k = 1,\ 2,\ 3.$$

Assumption (C) For all $\boldsymbol{\varphi} \in \Theta$, the Fisher information matrix $I(\boldsymbol{\varphi})$ is positive definite.

Assumption (D) There exists a function $M_{jk\ell}$ such that

$$\left|\frac{\partial^3 \log f_{\boldsymbol{\varphi}}(t,\ \delta,\ \delta^*)}{\partial\phi_j\partial\phi_k\partial\phi_\ell}\right| \le M_{jk\ell}(t,\ \delta,\ \delta^*),\quad \text{for all } (t,\ \delta,\ \delta^*) \in \chi,\quad \boldsymbol{\varphi} \in \omega,$$

where $m_{jk\ell} = E_{\boldsymbol{\varphi}^0}\{M_{jk\ell}(T,\ \Delta,\ \Delta^*)\} < \infty$ for $j,\ k,\ \ell = 1,\ 2,\ 3$.

Assumption (A) holds if one does not impose unusual constraints on Θ. Assumption (B) holds under a mild assumption on the interchangeability between differentiation and integration; see Supplementary Material for the proof. Assumption (B) implies

Table 3 Limits of the third-order derivatives of $\log f_{1,(\alpha,\gamma)}(t)$ and $\log \psi_{\boldsymbol{\varphi}}(1,\ t)$

$\log f_{1,(\alpha,\gamma)}(t)$	$t \to 0$	$t \to \infty$	$\log \psi_{\boldsymbol{\varphi}}(1,\ t)$	$t \to 0$	$t \to \infty$
$\log f_{1,(\alpha,\gamma)}(t)$	$t \to 0$	$t \to \infty$	$\log \psi_{\boldsymbol{\varphi}}(1,\ t)$	$t \to 0$	$t \to \infty$
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \alpha^3}$	$\frac{2}{\alpha^3}$	$\frac{2}{\alpha^3}$	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \alpha^3}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \alpha^2 \partial \beta}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \alpha^2 \partial \beta}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \alpha^2 \partial \gamma}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \alpha^2 \partial \gamma}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \beta^3}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \beta^3}$	0	$\frac{2}{\beta^3}$
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \alpha \partial \beta^2}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \alpha \partial \beta^2}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \beta^2 \partial \gamma}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \beta^2 \partial \gamma}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \gamma^3}$	$\frac{2}{\gamma^3}$	$\frac{2}{\gamma^3}$	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \gamma^3}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \alpha \partial \gamma^2}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \alpha \partial \gamma^2}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \beta \partial \gamma^2}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \beta \partial \gamma^2}$	0	0
$\frac{\partial^3 \log f_{1,(\alpha,\gamma)}(t)}{\partial \alpha \partial \beta \partial \gamma}$	0	0	$\frac{\partial^3 \log \psi_{\boldsymbol{\varphi}}(1,t)}{\partial \alpha \partial \beta \partial \gamma}$	0	0

$$I_{jk}(\boldsymbol{\varphi}) = E_{\boldsymbol{\varphi}}\left(-\frac{\partial^2 \log f_{\boldsymbol{\varphi}}(T,\ \Delta,\ \Delta^*)}{\partial \phi_j \partial \phi_k}\right), \quad j,\ k = 1,\ 2,\ 3. \tag{7}$$

Assumption (C) is hard to verify analytically. As an alternative, we check the positive definiteness of $\hat{I}(\hat{\boldsymbol{\varphi}}_n)$ that will be defined in Sect. 4.2. Such conditions always hold in our numerical analyses. Table 3 shows that all the derivative expressions in Assumption (D) are bounded. Assumption (D) holds under a simple assumption that will be given in Sect. 4.3.

Theorem 1 *If Assumptions* (A)–(D) *hold, then* (a) *Existence and consistency: With probability tending to 1 as $n \to \infty$, there exists the MLE $\hat{\boldsymbol{\varphi}}_n = (\hat{\alpha}_n,\ \hat{\beta}_n,\ \hat{\gamma}_n)$ such that $\hat{\boldsymbol{\varphi}}_n \xrightarrow{p} \boldsymbol{\varphi}^0$ as $n \to \infty$.* (b) *Asymptotic normality: $\sqrt{n}(\hat{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}^0) \xrightarrow{d} N_3(\mathbf{0},\ I(\boldsymbol{\varphi}^0)^{-1})$ as $n \to \infty$.*

The proof follows Theorem 6.5.1 of Lehmann and Casella (1998). However, Assumptions (A)–(D) have been modified for the competing risks setting in which the density $f_{\boldsymbol{\varphi}}(t,\ \delta,\ \delta^*)$ and the bound $M_{jk\ell}(t,\ \delta,\ \delta^*)$ are the mixture of continuous and discrete parts.

4.2 Standard error and confidence interval

We use Theorem 1 to obtain the standard error (SE) and the confidence interval (CI). Based on Eq. (7), we obtain an approximation to the Fisher information matrix

$$I(\boldsymbol{\varphi}^0) \approx \hat{I}(\hat{\boldsymbol{\varphi}}_n) \equiv -\frac{1}{n} \left. \frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^{\mathrm{T}}} \right|_{\boldsymbol{\varphi}=\hat{\boldsymbol{\varphi}}_n},$$

where $\hat{I}(\hat{\boldsymbol{\varphi}}_n)$ is available from the last step in Algorithm 1. For instance, the SE of $\hat{\alpha}_n$ is $SE(\hat{\alpha}_n) = [\hat{I}^{-1}(\hat{\boldsymbol{\varphi}}_n)]_\alpha$, where $[\ \ ]_\alpha$ is the diagonal element corresponding to α. By Theorem 1 and a log-transformation, the 95% CI of α^0 is $\hat{\alpha}_n \times \exp\{\pm 1.96 \times SE(\hat{\alpha}_n)/\hat{\alpha}_n\}$.

The delta method applies to a differentiable function $g : \mathrm{R}^3 \mapsto \mathrm{R}$ to obtain

$$SE\{g(\hat{\boldsymbol{\varphi}}_n)\} = \sqrt{\left. \left\{ \frac{\partial g(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} \right\}^{\mathrm{T}} \left\{ -\frac{\partial^2 \ell_n(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^{\mathrm{T}}} \right\}^{-1} \left\{ \frac{\partial g(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} \right\} \right|_{\boldsymbol{\varphi}=\hat{\boldsymbol{\varphi}}_n}}.$$

For example, the mean failure time for cause 1 failure is

$$\mu_X = g(\boldsymbol{\varphi}) = E_{\boldsymbol{\varphi}}(X) = \alpha B\left(\alpha + \frac{1}{\gamma},\ 1 - \frac{1}{\gamma}\right).$$

The SE of $\hat{\mu}_X = g(\hat{\boldsymbol{\varphi}}_n)$ is obtained from

$$\frac{\partial g(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = \left(\frac{\partial g(\boldsymbol{\varphi})}{\partial \alpha}, \frac{\partial g(\boldsymbol{\varphi})}{\partial \beta}, \frac{\partial g(\boldsymbol{\varphi})}{\partial \gamma} \right)^{\mathrm{T}},$$

$$\frac{\partial g(\boldsymbol{\varphi})}{\partial \alpha} = \alpha B\left(\alpha + \frac{1}{\gamma},\ 1 - \frac{1}{\gamma}\right) \left\{ \psi\left(\alpha + \frac{1}{\gamma}\right) - \psi(\alpha) \right\},$$

$$\frac{\partial g(\boldsymbol{\varphi})}{\partial \beta} = 0, \quad \text{and} \quad \frac{\partial g(\boldsymbol{\varphi})}{\partial \gamma} = \frac{\alpha}{\gamma^2} B\left(\alpha + \frac{1}{\gamma},\ 1 - \frac{1}{\gamma}\right) \left\{ \psi\left(1 - \frac{1}{\gamma}\right) - \psi\left(\alpha + \frac{1}{\gamma}\right) \right\},$$

where $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$ is the digamma function. Then the 95% CI of $\hat{\mu}_X$ is $\hat{\mu}_X \times \exp\{\pm 1.96 \times SE(\hat{\mu}_X)/\hat{\mu}_X\}$.

One might compute the SE and 95% CI using the bootstrap. However, it requires likelihood maximization for each resampling step, yielding a large number of randomizations and the NR iteration steps in Algorithm 1. Once we admit the asymptotic theory to hold, the SE and 95% CI can be obtained more quickly from the last step of Algorithm 1.

4.3 Simple sufficient condition

A simple sufficient condition to verify Assumption (D) is given below.

Table 4 Three censoring percentages under $(\alpha,\ \beta,\ \gamma) = (2,\ 2,\ 3)$

		$w = 6$	$w = 3$	$w = 2$
Cause 1 failure	$\Pr(X_i = T_i,\ Y_i > T_i) \times 100$	40%	30%	20%
Cause 2 failure	$\Pr(X_i > T_i,\ Y_i = T_i) \times 100$	40%	30%	20%
Censoring	$\Pr(X_i > T_i,\ Y_i > T_i) \times 100$	20%	40%	60%

Lemma 1 *Assumption* (D) *holds if* $\Theta = [u_\alpha,\ v_\alpha] \times [u_\beta,\ v_\beta] \times [u_\gamma,\ v_\gamma]$, *for some positive numbers* $0 < u_\alpha < v_\alpha$, $0 < u_\beta < v_\beta$ *and* $0 < u_\gamma < v_\gamma$.

The proof of Lemma 1 is given in Supplementary Material.

5 Simulation

This section evaluates the performance of the proposed methods by simulations.

5.1 Simulation design

To assess the performance of the maximum likelihood estimation, we generate data $(X_i,\ Y_i)$ for $i = 1,\ 2,\ \ldots,\ n$ from the generalized FGM copula model with Burr III margins, that is

$$F_{\boldsymbol{\varphi}}(x,\ y) = (1 + x^{-\gamma})^{-\alpha}(1 + y^{-\gamma})^{-\beta}[1 + \theta\,\{1 - (1 + x^{-\gamma})^{-\alpha p}\}^q\{1 - (1 + y^{-\gamma})^{-\beta p}\}^q],$$

where $p = 3$, $q = 2$, and $\theta = 0.7$ ($\tau = 0.284$). We choose the true parameters to be $(\alpha,\ \beta,\ \gamma) = (2,\ 2,\ 3)$, $(2,\ 4,\ 5)$, or $(3,\ 2,\ 7)$. We generate independent censoring time $C_i \sim U(0,\ w)$, where $w > 0$ is a constant yielding three different censoring percentage: light (20%), moderate (40%), and heavy (60%) (Table 4). Then, we obtain the data $(T_i,\ \delta_i,\ \delta_i^*)$, where $T_i = \min(X_i,\ Y_i,\ C_i)$, $\delta_i = \mathbf{I}(T_i = X_i)$, and $\delta_i^* = \mathbf{I}(T_i = Y_i)$ for $i = 1,\ 2,\ \ldots,\ n$.

Based on the data, we obtain the MLE $(\hat{\alpha},\ \hat{\beta},\ \hat{\gamma})$ along with their SEs using Algorithm 1 (Randomized NR algorithm). We count the number of NR iterations and the number of randomizations to assess the convergence speed of Algorithm 1. In addition, we obtain the MLEs $\hat{\mu}_X = E_{\hat{\boldsymbol{\varphi}}}(X)$ and $\hat{\mu}_Y = E_{\hat{\boldsymbol{\varphi}}}(Y)$, and their SEs. We compute the 95% CI to examine the coverage probability (CP) for each estimate. Our simulation results are based on 10,000 repetitions. Algorithm 1 ascertained the MLE in every repetition. Supplementary Material includes additional results on a sensitivity analysis under a copula misspecification.

To assess the performance of the goodness-of-fit tests, we generate data $(T_i,\ \delta_i,\ \delta_i^*)$ for $i = 1,\ 2,\ \ldots,\ n$ in a similar fashion under three different copulas (generalized FGM, Clayton, and independent) with two different margins (Burr III and exponential). The censoring percentages are controlled around 20% in all the cases. The dependence parameter in the generalized FGM and Clayton copulas are chosen to yield the same Kendall's tau, $\tau = 0.284$. The marginal parameters are $(\alpha,\ \beta,\ \gamma) = (1.5,\ 1,\ 1)$ for the Burr III margins and $(\lambda_1,\ \lambda_2) = (0.5,\ 1)$ for the exponential margins. For each repetition, we calculate the Cramér-von Mises type test statistics and test H_0^1 and H_0^2 at level $\alpha = 0.05$ with sample size $n = 200$ and bootstrap replications $B = 500$. Due to the high computational cost, the simulation results for the goodness-of-fit test are reported under 500 repetitions.

5.2 Simulation results

Table 5 gives the average values of $(\hat{\alpha},\ \hat{\beta},\ \hat{\gamma})$ and the mean square errors (MSEs). It shows that the estimates are almost unbiased and the MSEs decrease when the sample size increase. On the other hand, the MSEs increase as the censoring percentage gets high. Even under the heavy censoring cases, the estimates are still nearly unbiased for $n = 200$ and 300. Table 5 also reveals that Algorithm 1 converges quickly (4–7 iterations on average). In the case of $(\alpha,\ \beta,\ \gamma) = (3,\ 2,\ 7)$, randomizations on initial values are frequently required. Hence, the randomization scheme is necessary to stabilize the algorithm.

Table 6 gives the standard deviations (SDs) of the MLE $(\hat{\alpha},\ \hat{\beta},\ \hat{\gamma})$ and their average SEs. It reveals that the SDs of the estimates are all close to the average SEs. Also, the CPs of the 95% CIs are very close to the nominal 0.95 in all the cases. The results support our asymptotic theory of Sect. 4.

Table 7 presents the SDs of $\hat{\mu}_X = E_{\hat{\boldsymbol{\varphi}}}(X)$ and $\hat{\mu}_Y = E_{\hat{\boldsymbol{\varphi}}}(Y)$, and their average SEs. In all cases, the SDs are close to the average SEs. The CPs of the 95% CIs are also close to 0.95. The results show that the delta method and the asymptotic theory works properly.

Table 8 shows the performance of the goodness-of-fit tests. When the data are generated under the correct model (generalized FGM copula with the Burr III margins), the rejection rates of the tests agree with the nominal level 0.05. In addition, the mean of the test statistic S_k is very close to the mean of its bootstrap versions. This indicates that the bootstrap approximation to the null distribution of the test statistic works fairly well.

When data are generated from the generalized FGM copula (correct copula) with the exponential margins (incorrect margins), the overall test has strong power to reject the null hypothesis. Hence, if the copula is correctly specified, the wrong marginal models are easy to detect. Note, however, that the rejection of the overall test does not identify which one is incorrect between copula and margins. The test of the copula keeps the rejection rate below the nominal level. The results imply that the two tests lead to the correct model identification: the generalized FGM copula is accepted and the exponential margins are rejected.

Table 5 Simulation results on the MLE ($\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$) based on 10,000 repetitions

Parameters	Proportion	n	$E(\hat{\alpha})$	$E(\hat{\beta})$	$E(\hat{\gamma})$	$MSE(\hat{\alpha})$	$MSE(\hat{\beta})$	$MSE(\hat{\gamma})$	AI	AR	MR
$\alpha = 2$	$X_i = T_i$ (40%)	100	2.025	2.024	3.043	0.050	0.051	0.055	4.9	0	0
$\beta = 2$	$Y_i = T_i$ (40%)	200	2.012	2.012	3.020	0.025	0.024	0.025	4.9	0	0
$\gamma = 3$	$C_i = T_i$ (20%)	300	2.007	2.007	3.015	0.016	0.016	0.017	4.9	0	0
	$X_i = T_i$(30%)	100	2.027	2.028	3.055	0.061	0.060	0.078	4.9	0	0
	$Y_i = T_i$ (30%)	200	2.013	2.014	3.025	0.029	0.028	0.037	4.9	0	0
	$C_i = T_i$ (40%)	300	2.009	2.007	3.018	0.019	0.019	0.024	4.9	0	0
	$X_i = T_i$ (20%)	100	2.037	2.036	3.070	0.083	0.082	0.141	5.0	0	0
	$Y_i = T_i$ (20%)	200	2.015	2.017	3.035	0.038	0.037	0.065	5.0	0	0
	$C_i = T_i$ (60%)	300	2.009	2.009	3.025	0.025	0.025	0.043	5.0	0	0
$\alpha = 2$	$X_i = T_i$ (59%)	100	2.026	4.115	5.069	0.048	0.346	0.143	5.7	< 0.1	7
$\beta = 4$	$Y_i = T_i$ (21%)	200	2.014	4.057	5.034	0.023	0.152	0.067	5.7	< 0.1	5
$\gamma = 5$	$C_i = T_i$ (20%)	300	2.007	4.035	5.025	0.015	0.097	0.044	5.7	< 0.1	3
	$X_i = T_i$ (45%)	100	2.031	4.144	5.102	0.058	0.470	0.202	5.4	< 0.1	6
	$Y_i = T_i$ (15%)	200	2.015	4.068	5.047	0.027	0.197	0.093	5.4	< 0.1	2
	$C_i = T_i$ (40%)	300	2.012	4.044	5.029	0.018	0.122	0.059	5.3	< 0.1	3
	$X_i = T_i$ (30%)	100	2.038	4.205	5.133	0.075	0.735	0.318	5.2	< 0.1	5
	$Y_i = T_i$ (10%)	200	2.017	4.096	5.058	0.035	0.269	0.144	5.1	0	0
	$C_i = T_i$ (60%)	300	2.010	4.054	5.044	0.023	0.169	0.094	5.1	0	0

Table 5 continued

Parameters	Proportion	n	$E(\hat{\alpha})$	$E(\hat{\beta})$	$E(\hat{\gamma})$	$MSE(\hat{\alpha})$	$MSE(\hat{\beta})$	$MSE(\hat{\gamma})$	AI	AR	MR
$\alpha = 3$	$X_i = T_i$ (29%)	100	3.069	2.025	7.115	0.152	0.050	0.284	6.4	1.2	16
$\beta = 2$	$Y_i = T_i$ (51%)	200	3.028	2.014	7.050	0.068	0.024	0.130	6.2	1.4	12
$\gamma = 7$	$C_i = T_i$ (20%)	300	3.019	2.009	7.035	0.045	0.016	0.088	6.1	1.6	15
	$X_i = T_i$ (21%)	100	3.082	2.030	7.135	0.203	0.062	0.391	6.5	0.6	13
	$Y_i = T_i$ (39%)	200	3.037	2.015	7.067	0.088	0.030	0.180	6.5	0.6	11
	$C_i = T_i$ (40%)	300	3.026	2.011	7.044	0.058	0.019	0.116	6.5	0.6	11
	$X_i = T_i$ (14%)	100	3.107	2.039	7.201	0.306	0.087	0.652	6.3	0.1	10
	$Y_i = T_i$ (26%)	200	3.059	2.020	7.092	0.127	0.039	0.295	6.2	0.1	8
	$C_i = T_i$ (60%)	300	3.035	2.011	7.063	0.079	0.025	0.183	6.2	< 0.1	8

AI the average number of iterations until convergence, *AR* the average number of randomizations until convergence, *MR* the maximum number of randomizations until convergence

$MSE(\hat{\alpha}) = E(\hat{\alpha} - \alpha)^2$, $MSE(\hat{\beta}) = E(\hat{\beta} - \beta)^2$, $MSE(\hat{\gamma}) = E(\hat{\gamma} - \gamma)^2$

Table 6 Simulation results on the MLE ($\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$) based on 10,000 repetitions

Parameters	Proportion	n	$\hat{\alpha}$			$\hat{\beta}$			$\hat{\gamma}$		
			SD	SE	CP	SD	SE	CP	SD	SE	CP
$\alpha = 2$	$X_i = T_i$ (40%)	100	0.222	0.219	0.951	0.224	0.219	0.948	0.231	0.225	0.942
$\beta = 2$	$Y_i = T_i$ (40%)	200	0.156	0.154	0.950	0.154	0.154	0.953	0.158	0.158	0.950
$\gamma = 3$	$C_i = T_i$ (20%)	300	0.126	0.125	0.949	0.126	0.125	0.951	0.129	0.129	0.948
	$X_i = T_i$ (30%)	100	0.245	0.240	0.951	0.244	0.240	0.950	0.274	0.270	0.945
	$Y_i = T_i$ (30%)	200	0.170	0.168	0.951	0.168	0.168	0.952	0.190	0.189	0.946
	$C_i = T_i$ (40%)	300	0.137	0.136	0.948	0.138	0.136	0.949	0.153	0.154	0.951
	$X_i = T_i$ (20%)	100	0.285	0.275	0.947	0.284	0.275	0.946	0.369	0.361	0.942
	$Y_i = T_i$ (20%)	200	0.194	0.191	0.951	0.191	0.191	0.952	0.253	0.251	0.947
	$C_i = T_i$ (60%)	300	0.157	0.155	0.946	0.158	0.155	0.947	0.205	0.204	0.947
$\alpha = 2$	$X_i = T_i$ (59%)	100	0.216	0.213	0.948	0.577	0.550	0.952	0.371	0.361	0.940
$\beta = 4$	$Y_i = T_i$ (21%)	200	0.151	0.149	0.949	0.385	0.378	0.951	0.256	0.253	0.946
$\gamma = 5$	$C_i = T_i$ (20%)	300	0.122	0.121	0.950	0.309	0.305	0.949	0.208	0.206	0.946
	$X_i = T_i$ (45%)	100	0.239	0.234	0.948	0.670	0.622	0.947	0.438	0.424	0.942
	$Y_i = T_i$ (15%)	200	0.165	0.163	0.948	0.439	0.424	0.949	0.301	0.296	0.946
	$C_i = T_i$ (40%)	300	0.135	0.133	0.948	0.346	0.342	0.951	0.242	0.241	0.947
	$X_i = T_i$ (30%)	100	0.270	0.262	0.946	0.833	0.739	0.952	0.548	0.533	0.940
	$Y_i = T_i$ (10%)	200	0.186	0.182	0.948	0.510	0.491	0.949	0.376	0.370	0.947

Table 6 continued

Parameters	Proportion	n	$\hat{\alpha}$			$\hat{\beta}$			$\hat{\gamma}$		
			SD	*SE*	CP	*SD*	*SE*	CP	*SD*	*SE*	CP
	$C_i = T_i$ (60%)	300	0.150	0.148	0.948	0.407	0.393	0.945	0.303	0.301	0.947
$\alpha = 3$	$X_i = T_i$ (29%)	100	0.384	0.370	0.948	0.222	0.216	0.945	0.520	0.507	0.940
$\beta = 2$	$Y_i = T_i$ (51%)	200	0.260	0.256	0.950	0.153	0.152	0.947	0.357	0.355	0.948
$\gamma = 7$	$C_i = T_i$ (20%)	300	0.211	0.208	0.947	0.125	0.123	0.946	0.295	0.289	0.944
	$X_i = T_i$ (21%)	100	0.443	0.418	0.948	0.247	0.241	0.948	0.610	0.592	0.940
	$Y_i = T_i$ (39%)	200	0.295	0.288	0.948	0.172	0.168	0.949	0.419	0.414	0.946
	$C_i = T_i$ (40%)	300	0.238	0.233	0.947	0.138	0.137	0.950	0.338	0.337	0.946
	$X_i = T_i$ (14%)	100	0.542	0.494	0.948	0.292	0.278	0.946	0.782	0.748	0.936
	$Y_i = T_i$ (26%)	200	0.351	0.338	0.948	0.197	0.193	0.946	0.535	0.518	0.942
	$C_i = T_i$ (60%)	300	0.279	0.272	0.948	0.157	0.157	0.952	0.424	0.421	0.947

SD standard deviation of the estimator, *SE* average standard error, *CP* coverage probability of the 95% confidence interval

Table 7 Simulation results on the estimates $\hat{\mu}_X = E_{\hat{\varphi}}(X)$ and $\hat{\mu}_Y = E_{\hat{\varphi}}(Y)$ based on 10,000 repetitions

Parameters	Proportion	n	$\hat{\mu}_X$			$\hat{\mu}_Y$		
			SD	SE	CP	SD	SE	CP
$\alpha = 2$	$X_i = T_i$ (40%)	100	0.107	0.106	0.944	0.107	0.106	0.943
$\beta = 2$	$Y_i = T_i$ (40%)	200	0.076	0.075	0.948	0.075	0.075	0.945
$\gamma = 3$	$C_i = T_i$ (20%)	300	0.062	0.061	0.946	0.062	0.061	0.944
	$X_i = T_i$ (30%)	100	0.129	0.127	0.940	0.128	0.127	0.942
	$Y_i = T_i$ (30%)	200	0.091	0.090	0.948	0.090	0.090	0.946
	$C_i = T_i$ (40%)	300	0.073	0.073	0.948	0.074	0.073	0.945
	$X_i = T_i$ (20%)	100	0.186	0.177	0.936	0.183	0.177	0.938
	$Y_i = T_i$ (20%)	200	0.123	0.121	0.945	0.121	0.121	0.950
	$C_i = T_i$ (60%)	300	0.100	0.098	0.944	0.100	0.098	0.944
$\alpha = 2$	$X_i = T_i$ (59%)	100	0.043	0.042	0.944	0.059	0.058	0.942
$\beta = 4$	$Y_i = T_i$ (21%)	200	0.030	0.030	0.947	0.041	0.041	0.949
$\gamma = 5$	$C_i = T_i$ (20%)	300	0.024	0.024	0.946	0.034	0.033	0.946
	$X_i = T_i$ (45%)	100	0.048	0.048	0.946	0.069	0.068	0.937
	$Y_i = T_i$ (15%)	200	0.034	0.034	0.949	0.048	0.048	0.946
	$C_i = T_i$ (40%)	300	0.028	0.028	0.946	0.039	0.039	0.951
	$X_i = T_i$ (30%)	100	0.060	0.060	0.942	0.089	0.086	0.943
	$Y_i = T_i$ (10%)	200	0.043	0.042	0.946	0.061	0.061	0.945
	$C_i = T_i$ (60%)	300	0.035	0.034	0.944	0.050	0.049	0.942
$\alpha = 3$	$X_i = T_i$ (29%)	100	0.032	0.032	0.946	0.028	0.027	0.941
$\beta = 2$	$Y_i = T_i$ (51%)	200	0.023	0.023	0.951	0.019	0.019	0.947
$\gamma = 7$	$C_i = T_i$ (20%)	300	0.019	0.018	0.945	0.016	0.016	0.949
	$X_i = T_i$ (21%)	100	0.038	0.037	0.940	0.032	0.031	0.943
	$Y_i = T_i$ (39%)	200	0.027	0.026	0.945	0.022	0.022	0.944
	$C_i = T_i$ (40%)	300	0.022	0.022	0.945	0.018	0.018	0.949
	$X_i = T_i$ (14%)	100	0.048	0.046	0.937	0.039	0.039	0.940
	$Y_i = T_i$ (26%)	200	0.034	0.033	0.944	0.028	0.027	0.942
	$C_i = T_i$ (60%)	300	0.027	0.027	0.949	0.022	0.022	0.948

SD standard deviation of the estimator, *SE* average standard error, *CP* coverage probability of the 95% confidence interval

When data are generated from the Clayton (incorrect) or the independence (incorrect) copula with the Burr III (correct) margins, the overall test has modest power to reject null hypothesis. The power increases if the marginal distributions are also incorrect. The test of the copula provides modest power to reject the generalized FGM copula. Hence, the two tests lead to the correct identification of the copula: the generalized FGM copula is rejected and the overall model is rejected. However, its power is low possibly due to the non-identifiability issue of the competing risks data.

Table 8 Simulation results on the goodness-of-fit tests based on sample size $n = 200$ with 500 repetitions

Underlying model		H_0^1 (testing overall model):			H_0^2 (testing copula)		
Copula	Margins	Test statistic	Bootstrap statistic	Rejection rate at level 0.05	Test statistic	Bootstrap statistic	Rejection rate at level 0.05
GFGM (correct)	Burr III (correct)	0.046	0.048	0.040	0.034	0.034	0.042
	Exponential (incorrect)	0.175	0.049	0.796	0.022	0.023	0.036
Clayton (incorrect)	Burr III (correct)	0.068	0.050	0.152	0.051	0.033	0.224
	Exponential (incorrect)	0.109	0.051	0.464	0.027	0.024	0.124
Independent (incorrect)	Burr III (correct)	0.068	0.049	0.176	0.041	0.027	0.222
	Exponential (incorrect)	0.086	0.050	0.290	0.027	0.022	0.142

Test statistic = The mean of the Cramér–von Mises type test statistics S_k, $k = 1, 2$

Bootstrap statistic = The mean of the averaged bootstrap Cramér-von Mises type test statistics $\Sigma_{b=1}^{B} S_k^{(b)}/B$, $B = 500$, $k = 1, 2$

Table 9 Summary of the radio data ($n = 369$) from Mendenhall and Hader (1958)

	Confirmed failure ($\delta_i = 1$)	Unconfirmed failure ($\delta_i^* = 1$)	Censoring ($\delta_i = \delta_i^* = 0$)
The number of events (event rate %)	218 (59%)	107 (29%)	44 (12%)
Average (h)	$\bar{X} = 229.6147$	$\bar{Y} = 191.1963$	$C = 630$ (fixed)

6 Data analysis

We illustrate the proposed method by using the radio transmitter-receivers data from Mendenhall and Hader (1958). The dataset contains failure times (hours) of the ARC-1 VHF radio transmitter-receivers of a single commercial airline. ARC-1 is the name of the radios and VHF stands for Very High Frequency. The radios were removed from the aircraft for maintenance when they are assumed to be failed. However, it turns out that some radios sent for maintenance were actually not failed. Therefore, the failure can be divided into two modes: confirmed failure and unconfirmed failure. Failure times were censored at 630 h due to the policy of the airline to remove radios having operated for 630 h.

We denote the confirmed latent failure time as X_i and the unconfirmed latent failure time as Y_i for subject $i = 1, 2, \ldots, 369$. The censoring time is fixed at $C_i = 630$ for all i (Type I censoring). Observations consist of triplets $(T_i, \delta_i, \delta_i^*)$, $i = 1, 2, \ldots, 369$, where $T_i = \min(X_i, Y_i, C_i)$, $\delta_i = \mathbf{I}(T_i = X_i)$, and $\delta_i^* = \mathbf{I}(T_i = Y_i)$. The data are summarized in Table 9.

As mentioned in Sect. 3.1, we fix the copula parameters (p, q, θ) to perform likelihood-based inference. Here, we set $p = 3$ and $q = 2$ to allow for a wide range of Kendall's tau (Shih and Emura 2016). To improve the fit of the data, we introduce a location parameter η to the generalized FGM copula model with Burr III margins as

$$F_{(\boldsymbol{\varphi},\eta)}(x, y) = F_{1,(\alpha,\gamma,\eta)}(x)F_{2,(\beta,\gamma,\eta)}(y) \times [1 + \theta\, \{1 - F_{1,(\alpha,\gamma,\eta)}(x)^p\}^q \{1 - F_{2,(\beta,\gamma,\eta)}(y)^p\}^q],$$

where

$$F_{1,(\alpha,\gamma,\eta)}(x) = \{1 + (x - \eta)^{-\gamma}\}^{-\alpha} \quad \text{and} \quad F_{2,(\beta,\gamma,\eta)}(y) = \{1 + (y - \eta)^{-\gamma}\}^{-\beta}.$$

Estimation procedures in Sect. 3.1 can be applied to the transformed data $(T_i - \eta, \delta_i, \delta_i^*)$, $i = 1, 2, \ldots, 369$. Now, given η and θ, the log-likelihood in Sect. 3.1 is modified as

$$\ell_n(\boldsymbol{\varphi}, \eta, \theta) = \sum_{i=1}^{n} \delta_i \log f_{\boldsymbol{\varphi}}(1, T_i - \eta) + \sum_{i=1}^{n} \delta_i^* \log f_{\boldsymbol{\varphi}}(2, T_i - \eta) + \sum_{i=1}^{n} (1 - \delta_i - \delta_i^*) \log \bar{F}_{\boldsymbol{\varphi}}(T_i - \eta, T_i - \eta).$$

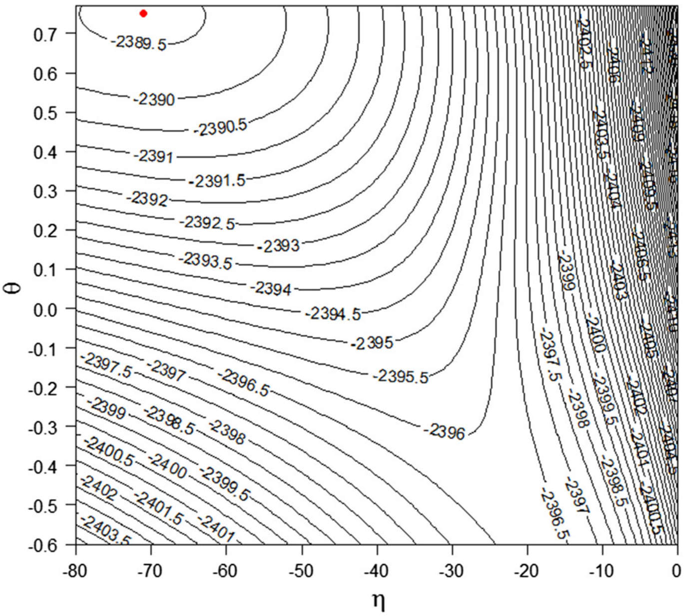

Fig. 1 The contour plot of the profile log-likelihood function. The point is drawn at $(\hat{\eta},\ \hat{\theta}) = (-71,\ 0.75)$ which reaches the maximum profile log-likelihood value

The profile log-likelihood of η and θ is defined as

$$\tilde{\ell}_n(\eta,\ \theta) = \ell_n(\hat{\boldsymbol{\varphi}}(\eta,\ \theta),\ \eta,\ \theta) = \max_{\boldsymbol{\varphi}} \ell_n(\boldsymbol{\varphi},\ \eta,\ \theta),$$

where $\hat{\boldsymbol{\varphi}}(\eta,\ \theta)$ is obtained by using Algorithm 1 (Randomized NR algorithm). We suggest applying a grid search method to obtain $(\hat{\eta},\ \hat{\theta}) = \arg\max_{(\eta,\theta)\in\Theta} \tilde{\ell}_n(\eta,\ \theta)$. Figure 1 gives the contour plot of the profile log-likelihood $\tilde{\ell}_n(\eta,\ \theta)$ within the admissible range $\theta \in (-0.604,\ 0.777)$ under $p = 3$ and $q = 2$. It reveals that the maximum value is attained at $(\hat{\eta},\ \hat{\theta}) = (-71,\ 0.75)$.

The positive dependence $\hat{\theta} = 0.75$ ($\hat{\tau} = 0.304$) seems to be reasonable since the radios with unconfirmed failure may still have a minor problem. The results imply that the unconfirmed failure time Y may predict the true (confirmed) failure time X.

We perform the proposed goodness-of-fit tests by using Algorithm 1. The results show that the test of the overall model is accepted ($S_1 = 0.0375$; P-value $= 0.502$) and the test of the generalized FGM copula is also accepted ($S_2 = 0.0075$; P-value $=$ 0.992). Therefore, there is no evidence against the generalized FGM copula model with the Burr III margins. Figure 2 shows the model-diagnostic plot that compares the three estimators of the sub-distribution functions (parametric, semi-parametric, and non-parametric). We see that the three estimators are close one another, implying that the model fits the data well.

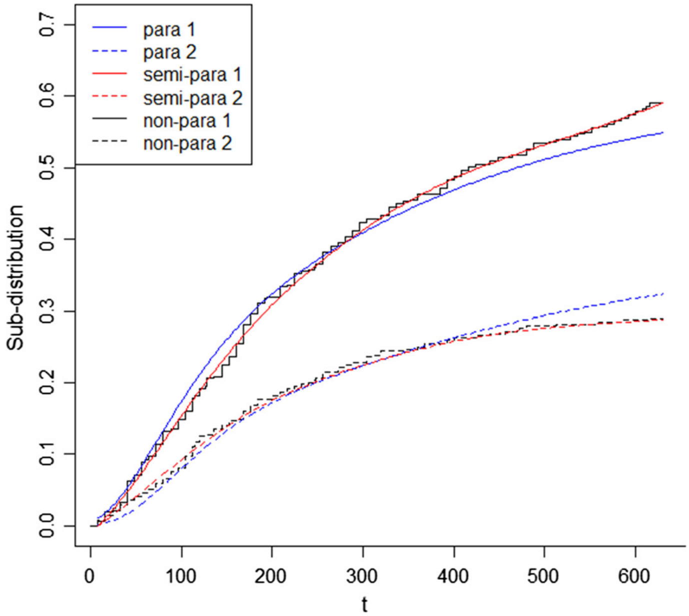

Fig. 2 Parametric (Burr III), semi-parametric (spline), and non-parametric estimators of the sub-distribution functions (causes 1 and 2) based on the radio data

Table 10 Parameter estimates based on the radio data from Mendenhall and Hader (1958)

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\mu}_X$	$\hat{\mu}_Y$
Estimate	1326.8	1835.3	1.298	938.8	1225.6
95% CI	(817.6, 2153.1)	(1132.4, 2974.4)	(1.211, 1.392)	(693.1, 1263.6)	(890.7, 1677.2)

The mean failure times are $\hat{\mu}_X = \hat{\alpha}B(\hat{\alpha}+1/\hat{\gamma},\ 1-1/\hat{\gamma})+\hat{\eta}$ and $\hat{\mu}_Y = \hat{\beta}B(\hat{\beta}+1/\hat{\gamma},\ 1-1/\hat{\gamma})+\hat{\eta}$, where $\hat{\eta} = -71$

Table 10 summarizes parameter estimates and their 95% CIs. Algorithm 1 converges quickly (in 6 iterations) without randomization. This implies that the initial value is appropriate. Figure 3 confirms that $(\hat{\alpha},\ \hat{\beta},\ \hat{\gamma})$ attains the maximum of the likelihood.

Figure 4 shows the fitted marginal density $f_j(t)$ and distribution function $F_j(t)$, $j = 1,\ 2$. It reveals that the frequency of confirmed failure $(f_1(t))$ is higher than the frequency of unconfirmed failure $(f_2(t))$ within the first 200 h. Beyond 200 h, the failure may be more likely due to non-fatal problems or human errors.

Figure 2 shows that the spline model fits better than the Burr III model. However, the spline model requires 10 parameters, and these parameters are hard to interpret. In addition, the spline model yields an improper distribution, i.e., the support is defined up to the largest observed failure time. The Burr III model provides an interpretable model with only 4 parameters (including the location parameter). Moreover, one can estimate the mean failure times under the Burr III model (Table 10), which are not

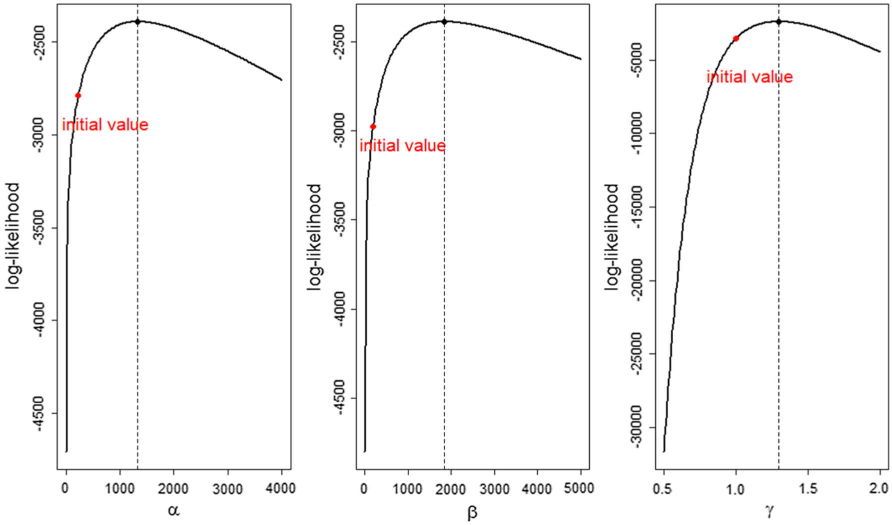

Fig. 3 Profile plots of the log-likelihood function based on the radio data. The vertical lines signify the MLEs $\hat{\alpha} = 1326.8$, $\hat{\beta} = 1835.3$, and $\hat{\gamma} = 1.298$

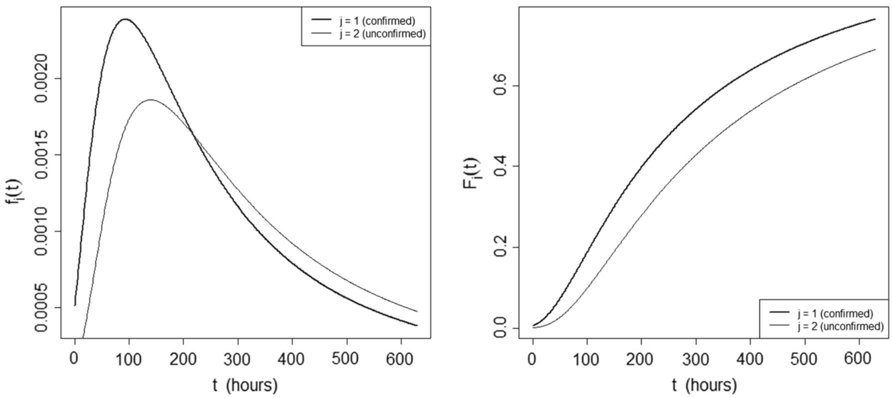

Fig. 4 The fitted densities and distribution functions based on the radio data

derived under the spline model. Hence, the Burr III model seems more appealing than the spline model in reliability analysis.

7 Discussion

There is a large body of literature on parametric inference procedures for latent failure time models with competing risks (Moeschberger 1974; Basu and Ghosh 1978; Crowder 2001; Escarela and Carrière 2003; Fan and Hsu 2015; Emura et al. 2015; Staplin et al. 2015; Hsu et al. 2016; Emura and Michimae 2017). Nevertheless, these parametric procedures appear to lack a competitive advantage over the semiparametric procedures in a few important aspects.

First, most parametric procedures employ the asymptotic theory of the MLE, but surprisingly, no paper seems to examine the regularity conditions for the theory. Hence, we still remain in ignorance on the validity of asymptotic inference in these inference procedures. On the other hand, the asymptotic theory is rigorously analyzed under the semi-parametric approaches (Rivest and Wells 2001; Braekers and Veraverbeke 2005; de Uña-Álvarez and Veraverbeke 2013, 2017). Our paper examined the regularity conditions for the proposed inference procedure under the class of the generalized FGM copula with the Burr III margins.

Second, the model-diagnostic procedures or goodness-of-fit tests are rarely discussed except for Escarela and Carrière (2003). These are important issues in parametric inference due to its strong reliance on model assumptions. In this paper, based on the explicit sub-distribution functions under the Burr III model, we have developed formal goodness-of-fit tests that refined the original idea of Escarela and Carrière (2003).

The location parameter η introduced in the data analysis (Sect. 6) is not included in the asymptotic theory (Sect. 4). The inclusion of η will make the log-likelihood non-differentiable since η is the lower support of the failure times. To further increase the flexibility of our model, it is possible to introduce a scale parameter σ to the Burr III marginal distribution to yield the four-parameter Burr III distribution considered by Lindsay et al. (1996). Concretely, one has

$$F_{1,(\alpha,\gamma,\eta,\sigma)}(x) = \left\{1 + \left(\frac{x-\eta}{\sigma}\right)^{-\gamma}\right\}^{-\alpha} \quad \text{and}$$

$$F_{2,(\beta,\gamma,\eta,\sigma)}(y) = \left\{1 + \left(\frac{y-\eta}{\sigma}\right)^{-\gamma}\right\}^{-\beta},$$

where $x,\ y > \eta, \alpha,\ \beta,\ \gamma,\ \sigma > 0$. Moreover, one may also consider different location and scale parameters for different marginal distributions.

Acknowledgements The authors thank the co-editor and two anonymous reviewers for their helpful comments that greatly improved the presentation of the paper. This work is supported by the research grants funded by the government of Taiwan (MOST 103-2118-M-008-MY2; MOST 105-2118-M-008-003-MY2).

References

Amini M, Jabbari H, Mohtashami Borzadaran GR (2011) Aspects of dependence in generalized Farlie–Gumbel–Morgenstern distributions. Commun Stat Simul Comput 40(8):1192–1205

Bairamov I, Bayramoglu K (2013) From the Huang–Kotz FGM distribution to Baker's bivariate distribution. J Multivar Anal 113:106–115

Bairamov I, Kotz S (2002) Dependence structure and symmetry of Huang–Kotz FGM distributions and their extensions. Metrika 56(1):55–72

Bakoyannis G, Touloumi G (2012) Practical methods for competing risks data: a review. Stat Methods Med Res 21(3):257–272

Basu AP, Ghosh JK (1978) Identifiability of the multinormal and other distributions under competing risks model. J Multivar Anal 8(3):413–429

Braekers R, Veraverbeke N (2005) A copula-graphic estimator for the conditional survival function under dependent censoring. Can J Stat 33(3):429–447
Burr IW (1942) Cumulative frequency functions. Ann Math Stat 13:215–232
Chen Y-H (2010) Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. J R Stat Soc B 72(2):235–251
Crowder MJ (2001) Classical competing risks. Chapman and Hall/CRC, Boca Raton
David HA, Moeschberger ML (1978) The theory of competing risks, vol 39. Griffin, London
de Uña-Álvarez J, Veraverbeke N (2013) Generalized copula-graphic estimator. Test 22(2):343–360
de Uña-Álvarez J, Veraverbeke N (2017) Copula-graphic estimation with left-truncated and right-censored data. Statistics 51(2):387–403
Domma F, Giordano S (2013) A copula-based approach to account for dependence in stress-strength models. Stat Pap 54(3):807–826
Domma F, Giordano S (2016) Concomitants of m-generalized order statistics from generalized Farlie-Gumbel-Morgenstern distribution family. J Comput Appl Math 294:413–435
Durante F, Sempi C (2015) Principles of copula theory. CRC Press, Boca Raton
Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, London
Emura T (2017) joint.Cox: Penalized likelihood estimation and dynamic prediction under the joint frailty-copula models between tumour progression and death for meta-analysis, CRAN, R package version 2.13
Emura T, Chen Y-H (2016) Gene selection for survival data under dependent censoring: a copula-based approach. Stat Methods Med Res 25(6):2840–2857
Emura T, Chen YH (2018) Analysis of Survival Data with Dependent Censoring. Copula-Based Approaches. JSS Research Series in Statistics, Springer, Singapore
Emura T, Michimae H (2017) A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae. Environ Ecol Stat 24(1):151–173
Emura T, Nakatochi M, Murotani K, Rondeau V (2015) A joint frailty-copula model between tumor progression and death for meta-analysis. Stat Methods Med Res. https://doi.org/10.1177/0962280215604510
Escarela G, Carrière JF (2003) Fitting competing risks with an assumed copula. Stat Methods Med Res 12(4):333–349
Eyraud H (1936) Les principes de la mesure des correlations. Ann Univ Lyon III Ser Sect A 1:30–47
Fan T-H, Hsu T-M (2015) Statistical inference of a two-component series system with correlated log-normal lifetime distribution under multiple type-I censoring. IEEE Trans Reliab 64(1):376–385
Gumbel EJ (1960) Bivariate exponential distributions. J Am Stat Assoc 55(292):698–707
Heckman JJ, Honore BE (1989) The identifiability of the competing risks models. Biometrika 76:325–330
Hsu T-M, Emura T, Fan T-H (2016) Reliability inference for a copula-based series system life test under multiple type-I censoring. IEEE Trans Reliab 65(1):1069–1080
Hu Y-H, Emura T (2015) Maximum likelihood estimation for a special exponential family under random double-truncation. Comput Stat 30(4):1199–1229
Huang JS, Kotz S (1999) Modifications of the Farlie–Gumbel–Morgenstern distributions. A tough hill to climb. Metrika 49:135–145
Knight K (2000) Mathematical statistics. Chapman and Hall, Boca Raton
Lawless JF (2003) Statistical models and methods for lifetime data, 2nd edn. Wiley, Hoboken
Lehmann EL, Casella G (1998) Theory of point estimation, 2nd edn. Springer, New York
Lindsay SR, Wood GR, Woollons RC (1996) Modelling the diameter distribution of forest stands using the Burr distribution. J Appl Stat 23(6):609–620
Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York
Mendenhall W, Hader RJ (1958) Estimation of parameters of mixed exponential distributed failure time distribution from censored life test data. Biometrika 45(3–4):504–520
Moeschberger ML (1974) Life tests under dependent competing causes of failure. Technometrics 16(1):39–47
Mokhlis NA (2005) Reliability of a stress-strength model with Burr type III distributions. Commun Stat Theory Methods 34(7):1643–1657
Morgenstern D (1956) Einfache Beispiele zweidimensionaler Verteilungen. Mitteilungsblatt für Mathematische Statistik 8:234–235
Navarro J (2016) Distribution-free comparison of residual lifetimes of coherent systems based on copula properties. Stat Pap. https://doi.org/10.1007/s00362-016-0789-0

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York
Ramsay JO (1988) Monotone regression splines in action. Stat Sci 3(4):425–461
Rivest LP, Wells MT (2001) A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. J Multivar Anal 79:138–155
Shao Q (2000) Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach. Environmetrics 11:583–595
Sherrick BJ, Garcia P, Tirupattur V (1996) Recovering probabilistic information form option markets: test of distributional assumptions. J Futures Mark 16(5):545–560
Shih J-H (2018) R GFGM.copula: Generalized Farlie–Gumbel–Morgenstern Copula, CRAN: version 1.0.2
Shih J-H, Emura T (2016) Bivariate dependence measures and bivariate competing risks models under the generalized FGM copula. Stat Pap. https://doi.org/10.1007/s00362-016-0865-5
Staplin ND, Kimber AC, Collett D, Roderick PJ (2015) "Dependent censoring in piecewise exponential survival models. Stat Methods Med Res 24(3):325–341
Tsiatis A (1975) A nonidentifiability aspect of the problem of competing risks. Proc Natl Acad Sci USA 72(1):20–22
Zheng M, Klein JP (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika 82(1):127–138