I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2012/32

On the identifiability of copulas in bivariate competing risks models

SCHWARZ, M., JONGBLOED, G. and I. VAN KEILEGOM

On the identifiability of copulas in bivariate competing risks models

Maik Schwarz* Geurt Jongbloed⋆ Ingrid Van Keilegom*

November 7, 2012

Abstract. In competing risks models, the joint distribution of the event times is not identifiable even when the margins are fully known, which has been referred to as the "identifiability crisis in competing risks analysis" (Crowder, 1991). We model the dependence between the event times by an unknown copula and show that identification is actually possible within many frequently used families of copulas. The result is then extended to the case where one margin is unknown.

MSC 2010 subject classifications: Primary 62N01, secondary 62N99

Keywords: Copulas, competing risks, identification, dependent censoring, bivariate distribution

1. Introduction

The theory of competing risks is concerned with the analysis of multiple possible causes of a certain event ("risk") in a system of interest. As an example, consider an animal experiment in which mice die either from a disease (event time X) or from the side effects of some treatment (event time Y). In this setting, one observes only the minimum of the two event times, corresponding to the time of death, and a variable indicating the cause of death, that is

$$Z := X \wedge Y \quad \text{and} \quad \Delta := \mathbf{1}_{X \leq Y}. \tag{1.1}$$

In this paper, we investigate under which conditions the joint distribution of (X, Y) can be identified based on the observations. Of course, in practice one could consider many more potential risks even in this simple example, but in this paper we treat the bivariate case exclusively. The focus lies in particular on the case where there

*Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Belgium. {maik.schwarz|ingrid.vankeilegom}@uclouvain.be

⋆Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands. G.Jongbloed@tudelft.nl

is dependence between the two risks X and Y, which is modelled by a bivariate copula $C := C_{XY}$. Indeed, the assumption of independence between X and Y seems unrealistic in a practical context. In the animal experiment example, one might suspect a positive correlation between the two risks, because it is conceivable that the vulnerability to the treatment's side effects depends on the course of the disease.

The identifiability of joint distributions in competing risks models is not a recent topic. Already in the 1950s, Cox (1959) pointed out an identifiability problem in the bivariate case for independent risks, and Tsiatis (1975) showed in the general multivariate case that the joint distribution of the failure times cannot be identified by their minimum. Crowder (1991) showed that "the situation is even worse than previously described" – even when the marginal laws of X and Y are known, the joint distribution function is not identifiable. This is a very undesirable property of the model, and so Crowder called out the "identifiability crisis in competing risks analysis".

Tsiatis' observations have been the starting point for investigations on conditions or modifications of the model which allow for identification and estimation of the event time distribution, and a variety of such models has been studied over the past decades. In order to obtain identifiability of the joint distribution of X and Y, one has to restrict the class of possible models. For example, one can exclude the independent case and restrict the class of admissible dependence structures. Basu and Ghosh (1978) follow this approach and show identifiability of the bivariate normal and the bivariate exponential distributions introduced in Marshall and Olkin (1967) and Gumbel (1960). Instead of fixing the precise dependence structure of the joint distribution in advance, Slud and Rubinstein (1983) suppose that a certain hazard ratio involving the event and the censoring time is known. Under such assumptions, they derive pointwise bounds on the marginal survivor function. Ebrahimi et al. (2003) use still a different condition, involving a partial derivative of the conditional survivor function of the event time given that the censoring time is larger than a given threshold.

Another approach which has been the topic of many research articles consists in modelling the dependence structure between the event and the censoring time using a copula. In view of Sklar's theorem, this kind of model allows for flexibility in the modelling of the dependence structure without affecting hypotheses on the margins. The identifiability of the copula and the margins can then be treated in two subsequent steps.

Zheng and Klein (1995) and Rivest and Wells (2001) suppose that the event and censoring times are dependent via some known copula that is nowhere constant. They show that their marginal distributions are identifiable if their support is $(0, \infty)$ and develop the "copula graphic estimator". Klein and Moeschberger (1988) also work under the hypothesis of a known copula; more specifically, they choose a Clayton copula with known parameter.

Of course, supposing the copula to be known is quite as unrealistic as supposing the risks X and Y to be independent, but in view of the results in Crowder (1991) cited above, identification is impossible if the copula is completely unknown. It is therefore natural to ask if the copula can be identified within certain (parametric) classes. It is the scope of the present work to answer this question. To this end, we proceed as follows.

In Section 2, we assume both of the marginal distributions to be entirely known and concentrate on the identification of the copula C which describes the dependence between the risks X and Y. Although the practical relevance of this setting is limited, its investigation is very instructive. We develop assumptions on the family of admissible copulas such that identification is possible based on the joint density of the observations, and we show that various parametric classes of copulas do actually satisfy such conditions. In particular, making use of a recent result by Wysocki (2012), we prove that many well-known classes of Archimedean copulas are identifiable. Certain classes of asymmetric copulas turn out to be identifiable as well. We also give examples of symmetric and asymmetric classes of copulas which cannot be identified.

In Section 3, we treat the more realistic case where F_X is completely unknown. Not surprisingly, additional model assumptions are needed in order to obtain identifiability in this more general setting. In particular, we assume that Δ and Z are stochastically independent. In the case where the copula C_{XY} is known, Braekers and Veraverbeke (2008) develop an estimator of F_X even when F_Y is unknown. Contrary to this, we show that even if the copula is unknown, F_X can be identified if F_Y is known. It is obviously impossible to consider all copulas. However, we will see that the same subclasses of Archimedean copulas as in Section 2 allow for identification. We will elaborate on the advantages and limitations of the model in Section 3 and discuss its relation to the classical Koziol-Green model.

The simultaneous identification of F_X and the copula in more general models remains an open question. Though the techniques developed in the present work will hopefully prove helpful in its solution, the identifiability crisis is not over yet and requires further investigation, as we finally discuss in Section 4.

2. Known margins

Throughout this section, suppose that the distributions F_X and F_Y of event and censoring time X and Y are known. The observations $(Z, \Delta) \in \mathbb{R} \times \{0, 1\}$ are given by (1.1). Assuming that the copula C is absolutely continuous with respect to the Lebesgue measure on the unit square with density c in the sense that

$$C(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} c(u, v)\, du dv,$$

we can write

$$\mathbf{P}[X \leq x, Y \leq y] = C(F_X(x), F_Y(y)) = \int_0^{F_X(x)} \int_0^{F_Y(y)} c(u,v)\,du\,dv$$

Consequently,

$$\mathbf{P}[Z > z, \Delta = 1] = \mathbf{P}[Y > X > z] = \int_{x=F_X(z)}^1 \int_{y=F_Y(F_X^{-1}(x))}^1 c(x,y)\,dy\,dx$$

and we obtain for the joint density h of (Z, Δ) that

$$\begin{aligned} h(z,0) &= f_Y(z) \int_{x=F_X(z)}^1 c(x, F_Y(z))dx = f_Y(z) - \frac{\partial}{\partial y} C(F_X(z), F_Y(y))\Big|_{y=z} \\ h(z,1) &= f_X(z) \int_{y=F_Y(z)}^1 c(F_X(z), y)dy = f_X(z) - \frac{\partial}{\partial x} C(F_X(x), F_Y(z))\Big|_{x=z}. \end{aligned} \tag{2.1}$$

Note that this implies the basic equality

$$C(F_X(z), F_Y(z)) = F_X(z) + F_Y(z) - H(z), \tag{2.2}$$

where H denotes the distribution function of Z. From observation (2.1), one sees that two copulas can be distinguished based on the distribution of (Z, Δ) if and only if either of their partial derivatives do not coincide on the curve

$$\Psi = (F_X(t), F_Y(t))_{t \in \mathbb{R}}. \tag{2.3}$$

This proves the following result.

Theorem 2.1 *Let X and Y be random variables with differentiable distribution functions F_X and F_Y, respectively, jointly distributed according to an unknown copula belonging to a class $\mathcal{C}$ of copulas having a density, and let (Z, Δ) be the observable random variables defined in (1.1). The class $\mathcal{C}$ is identifiable based on the joint distribution of (Z, Δ) if and only if for any two different copulas $C, C' \in \mathcal{C}$ there exists $z \in \mathbb{R}$ such that*

$$\frac{\partial}{\partial x} C(F_X(x), F_Y(z))\Big|_{x=z} \neq \frac{\partial}{\partial x} C'(F_X(x), F_Y(z))\Big|_{x=z}$$
$$\text{or} \quad \frac{\partial}{\partial y} C(F_X(z), F_Y(y))\Big|_{y=z} \neq \frac{\partial}{\partial y} C'(F_X(z), F_Y(y))\Big|_{y=z}.$$

This theorem provides us with a necessary and sufficient identifiability criterion for arbitrary classes of absolutely continuous copulas and margins admitting densities. In practice, this criterion will be difficult to verify for a given class, though. As noted, the class of all copulas is far too big to be identifiable as has been shown by Crowder (1991). In the remainder of the present section we will therefore restrict our attention to more particular classes of copulas and develop identification criteria that are easier to apply.

2.1. Symmetric copulas

It follows from representation (2.1) that whenever $F_X = F_Y$ and in addition both first partial derivatives of C with respect to x and y coincide along the curve Ψ defined in (2.3), the joint density h does not depend on d, which happens for instance when the copula is symmetric. In this case, the curve Ψ is the diagonal of the unit square, and the density of Z takes the simpler form

$$h(z) = 2f_Y(z) - 2\frac{\partial}{\partial y} C(F_Y(z), F_Y(y))|_{y=z} \tag{2.4}$$

or, equivalently,

$$\delta_C(F_Y(z)) := C(F_Y(z), F_Y(z)) = 2F_Y(z) - H(z), \tag{2.5}$$

where H denotes the distribution function of Z and where δ_C is called the diagonal section of C.

Theorem 2.2 *Suppose that the marginal distributions F_X and F_Y are arbitrary but known. Then, the class of all symmetric copulas is not identifiable.*

Proof. Consider first the case $F_X = F_Y$. Then the curve Ψ defined in (2.3) is the diagonal of the unit square. In order to show the non-identifiability of the class of symmetric copulas, we construct two distinct symmetric copulas which coincide on a strip of positive width around the diagonal and which hence yield the same distribution H. To this end, define first functions from $[0,1]^2$ to $\{0, \pm 1\}$ according to

$$q^{\eta}_{(x,y)}(s,t) := \mathrm{sgn}[(s-x)(t-y)]\, \mathbf{1}\{[(|s-x| \vee |t-y|) < \eta]\} + \mathrm{sgn}[(s-y)(t-x)]\, \mathbf{1}\{[(|s-y| \vee |t-x|) < \eta]\}.$$

The support of the function $q^{\eta}_{(x,y)}$ consists of two squares of side length 2η centred in (x,y) and (y,x), respectively. These squares in turn consist of four smaller squares on which $q^{\eta}_{(x,y)}$ is constant $+1$ or -1. Now let C be a symmetric copula admitting a density c and let $x, y, \eta, \varepsilon \in (0,1)$ be such that $\mathrm{supp}(q^{\eta}_{(x,y)})$ is in $[0,1]^2$ and does not intersect with a neighbourhood of the diagonal as illustrated in Figure 1 and $c(s,t) > \varepsilon$ for $(s,t) \in \mathrm{supp}(q^{\eta}_{(x,y)})$. It is then easy to see that the function $c + \varepsilon q^{\eta}_{(x,y)}$ is a density yielding a symmetric copula C' which coincides with C on a band of positive width around the diagonal. Thus, the two different symmetric copulas C and C' give rise to the same distribution H according to (2.1), showing that the class of symmetric copulas is not identifiable.

The general case $F_X \neq F_Y$ is straightforward applying the appropriate transformations. □

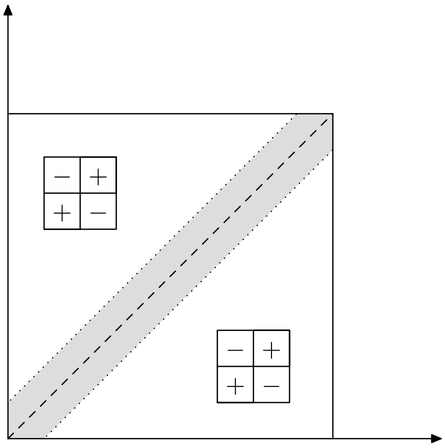

Figure 1: The support of the piecewise constant function $q^{\eta}_{(x,y)}$ consists of two squares that must not intersect with a neighbourhood of the diagonal

2.2. Archimedean copulas

We have seen that the class of all symmetric copulas is not identifiable. In this section, we restrict our attention further and consider only Archimedean copulas, that is copulas of the form

$$C(x, y) = \varphi^{[-1]}(\varphi(x) + \varphi(y))$$

for some generator function $\varphi \in \Omega$, where Ω denotes the set of decreasing convex functions from $[0, 1]$ to $[0, \infty]$ with $\varphi(1) = 0$. The pseudo-inverse $\varphi^{[-1]}$ of φ is given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{if } 0 \le t \le \varphi(0) \\ 0 & \text{if } t > \varphi(0). \end{cases}$$

In the case where φ is unbounded, the associated copula is called *strict* and we have $\varphi^{[-1]} = \varphi^{-1}$. Because of their nice properties, Archimedean copulas are very commonly used in the modelling of dependence and are therefore an interesting family of copulas to consider.

Recent work by Wysocki (2012) on the construction of copulas from diagonal sections allows us to state an identification result for a subfamily of the class of Archimedean copulas in the special case where the margins are equal.

Theorem 2.3 *Suppose that the marginal distributions F_X and F_Y are arbitrary but known. If $F_X = F_Y$, then the subclass of strict Archimedean copulas generated by*

$$\Omega_1 := \{\varphi \in \Omega \mid \varphi(0) = \infty,\ \lim_{t \uparrow 1} \varphi'(t) < 0\}$$

is identifiable.

Proof. The generator of an Archimedean copula being uniquely determined up to a multiplicative constant, we may assume that $\lim_{t\uparrow 1} \varphi'(t) = -1$ without loss of generality. Lemma 1 in Wysocki (2012) states that under this assumption, the diagonal sections of two Archimedean copulas coincide (up to a multiplicative constant) if and only if their generators do so, that is, if the copulas are the same. As by virtue of (2.5), the diagonal section δ_C of the copula can be identified when the margins are equal, this implies the identifiability of the whole copula. □

Theorem 2.3 implies identifiability for several well-known classes of Archimedean copulas when both margins are equal on the unit interval. Important examples of such classes can be found in Nelsen (2006, Table 4.1); in particular, the Frank copulas are identifiable. It is noteworthy that obviously even the union of all the identified subclasses of Ω_1 is identifiable. But the result is restricted to the special case of equal margins and to families of strict copulas. For example, the class of strict Clayton copulas is identifiable by virtue of the theorem, but some Clayton copulas are not strict ($\theta < 0$). Also note that the theorem cannot be applied to every family of strict Archimedean copulas. For instance, the result does not apply to the Gumbel family, because $\varphi'_\theta(1) = 0$ for all $\theta > 1$.

We have seen that Theorem 2.3 allows for identification within quite a large class of strict copulas, but only in the special case of equal margins. We still need a more general criterion for classes of strict copulas which are not covered by the theorem as well as for nonstrict copulas and general margins.

Before stating the next result, let us recall that the level sets of a copula C are given by $\{(u, v) \in [0, 1]^2 \mid C(u, v) = t\}$. For an Archimedean copula and for $t > 0$, this level set consists of the points on the level curve $\varphi(u) + \varphi(v) = \varphi(t)$. For $t = 0$, this curve is called zero curve; it is the boundary of the copula's zero set. All level curves of an Archimedean copula are convex.

Theorem 2.4 *Suppose that the marginal distributions F_X and F_Y are arbitrary but known. Let $\Phi := \{\varphi_\theta \mid \theta \in \Theta \subset \mathbb{R}\} \subset \Omega$ be a family of generators and write C_θ for the Archimedean copula generated by φ_θ. For $z \in [0, 1]$ denote by $\zeta^\theta(z) = (\zeta_x^\theta(z), \zeta_y^\theta(z))$ the intersection point of the copula C_θ's level curve of level z with the curve Ψ given in (2.3). If for some $z \in [0, 1]$ at least one of the two coordinates of $\zeta^\theta(z)$ is strictly monotonic as a function from Θ to $[0, 1]$, then the class of copulas generated by Φ is identifiable.*

Proof. Suppose without loss of generality that $\zeta_x^\theta(z)$ is strictly monotonic in θ. The distribution function H of (Z, Δ) is obviously identifiable. It is thus sufficient to show that the true parameter θ can be computed based on the knowledge of H, which we therefore suppose to be known. Then, by virtue of (2.2) and using that both margins are known, the copula can be reconstructed completely along the path Ψ. Consequently, the coordinate $\zeta_x^\theta(z)$ can be computed in this case for any choice of z.

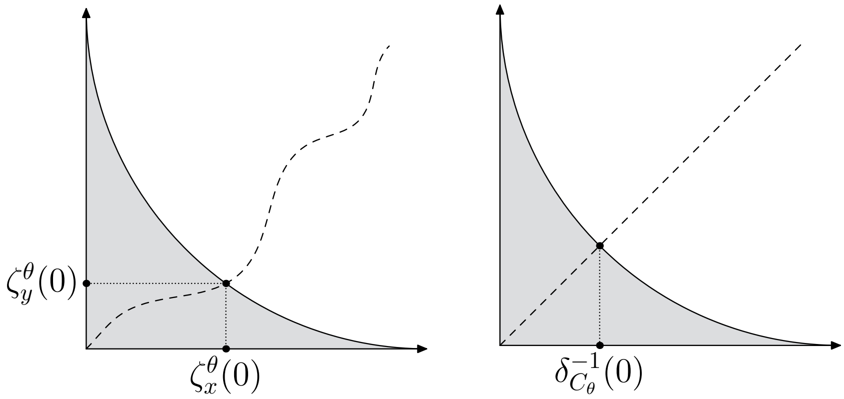

Figure 2: An Archimedean copula's zero curve and its intersection with the curve $\Psi = (F_X(t), F_Y(t))_{t\in\mathbb{R}}$ (on the left), and the special case $F_X = F_Y$ (on the right). The copula's zero set is shaded in grey.

(See Figure 2 on the left for an illustration of the case $z = 0$.) The strict monotonicity in θ ensures that $\zeta_x^\theta(z)$ uniquely characterises θ and thus C_θ. □

Example 2.5 Let us consider the special case where the marginal distributions F_X and F_Y are the same and strictly monotonic on their whole support. Let C be an Archimedean copula with generator $\varphi \in \Omega$. It is easily seen that we have, for $z \in (0, 1]$,

$$\begin{aligned}\delta_C(F_X(x)) = C(F_X(x), F_X(x)) = z \iff \varphi^{[-1]}(2\varphi(F_X(x))) = z \\ \iff \varphi(F_X(x)) = \frac{1}{2}\varphi(z) \iff F_X(x) = \varphi^{-1}\Big(\frac{1}{2}\varphi(z)\Big) =: \delta_C^{-1}(z).\end{aligned}$$

For $z = 0$, the second equivalence is not true in general. Nevertheless, the same definition of δ_C^{-1} can be used for $z = 0$ as well, with the convention that $\delta_C^{-1}(0) = \sup\{x \in \mathbb{R} \mid \delta_C(x) = 0\}$. The intersection point $\zeta(z)$ of the copula's level curve of level z with the diagonal can be expressed in terms of the inverse diagonal section, namely $\zeta(z) = (\delta_C^{-1}(z), \delta_C^{-1}(z))$. A parametric class of Archimedean copulas $\{C_\theta\}$ is thus identifiable if for some z, the inverse diagonal section $\delta_{C_\theta}^{-1}(z)$ is strictly monotonic as a function of θ.

Choosing $z = 1/2$, this condition can be applied to the Gumbel family. Recall that the Gumbel family could not be treated with Theorem 2.3. If φ is bounded (i.e. if C is nonstrict), $\delta_C^{-1}(0)$ is strictly positive and corresponds to the intersection of the diagonal with the boundary of the copula's zero set, see Figure 2 on the right. Supposing that φ belongs to a parametric class, knowledge of $\delta_C^{-1}(0)$ is often sufficient in order to characterise C within the class. Within the whole Archimedean class, though, there exist different copulas having the same zero set (cf. Nelsen, 2006, p. 132), such that this criterion fails.

Theorem 2.4 has a corollary that provides us with an identifiability criterion in terms of the generator.

Corollary 2.6 *Let $\Phi := \{\varphi_\theta \mid \theta \in \Theta \subset \mathbb{R}\} \subset \Omega$ be a family of bounded generators such that $\theta \mapsto \varphi_\theta(t)/\varphi_\theta(0)$ is strictly monotonic in θ for every $t \in (0,1)$. Suppose that the marginal distributions F_X and F_Y are arbitrary but known. Then, the copulas generated by the class Φ are identifiable.*

Proof. Recall that a constant times a generator results in the same Archimedean copula. Thus, without loss of generality, we can consider the following class of generators instead of Φ which generates the same copulas: $\tilde{\Phi} = \{\tilde{\varphi}_\theta := \varphi_\theta/\varphi_\theta(0) \mid \theta \in \Theta \subset \mathbb{R}\} \subset \Omega$. Note that by construction $\tilde{\varphi}_\theta(0) = 1$ for all θ and $\tilde{\varphi}_\theta(t)$ is strictly monotonic as a function of θ for all $t \in (0,1)$ (say strictly increasing, without loss of generality).

Consider the intersection point ζ^θ of the copula's zero curve with the curve Ψ defined in (2.3). This point can be identified by means of the distribution function H (cf. Theorem 2.4). We show below that for $\theta' > \theta$, the zero curve of C_θ lies strictly above the one of $C_{\theta'}$ on $(0,1)$. This implies that two different choices of θ always result in two different intersection points ζ and thus completes the proof.

The zero curve of the copula generated by $\tilde{\varphi}$ is given by $t \mapsto \tilde{\varphi}^{-1}(1 - \tilde{\varphi}(t))$ for $t \in [0,1]$. Let $t \in (0,1)$. We have that (using Lemma A.1 in the penultimate step)

$$\begin{aligned}
& \tilde{\varphi}_{\theta'}(t) > \tilde{\varphi}_\theta(t) \\
\Longleftrightarrow \quad & \tilde{\varphi}_{\theta'}(t) > \tilde{\varphi}_\theta(t) - \underbrace{(\tilde{\varphi}_\theta(0) - \tilde{\varphi}_{\theta'}(0))}_{=(1-1)=0} \\
\Longleftrightarrow \quad & \tilde{\varphi}_\theta(0) - \tilde{\varphi}_\theta(t) > \tilde{\varphi}_{\theta'}(\tilde{\varphi}_{\theta'}^{-1}(\tilde{\varphi}_{\theta'}(0) - \tilde{\varphi}_{\theta'}(t))) \\
\Longrightarrow \quad & \tilde{\varphi}_\theta(0) - \tilde{\varphi}_\theta(t) > \tilde{\varphi}_\theta(\tilde{\varphi}_{\theta'}^{-1}(\tilde{\varphi}_{\theta'}(0) - \tilde{\varphi}_{\theta'}(t))) \\
\Longleftrightarrow \quad & \tilde{\varphi}_\theta^{-1}(\tilde{\varphi}_\theta(0) - \tilde{\varphi}_\theta(t)) < \tilde{\varphi}_{\theta'}^{-1}(\tilde{\varphi}_{\theta'}(0) - \tilde{\varphi}_{\theta'}(t)).
\end{aligned}$$

This completes the proof. □

Example 2.7 The class (4.2.2) from Nelsen (2006), which is given by $\Phi := \{\varphi_\theta(t) = (1-t)^\theta \mid \theta \in [1,\infty)\}$, is identifiable by virtue of Corollary 2.6, because $\varphi_\theta(0) = 1$ for all θ, and $\varphi_\theta(t)$ is strictly decreasing in θ for all $t \in (0,1)$. In fact, Corollary 2.6 can be applied successfully to all classes of bounded generators listed in Nelsen (2006, Table 4.1). As for most of the remaining classes (strict or not), Theorem 2.4 can be applied successfully. The required computations become rather unwieldy, though.

2.3. Asymmetric copulas

So far we have restricted our attention to families of symmetric copulas, which contain some of the most widely used classes. In this section, we discuss briefly some families of asymmetric copulas. Let us consider a copula density of the form $c(x,y) = \widetilde{c}(x-y)$ with a 1-periodic function $\widetilde{c} : \mathbb{R} \to [0,\infty)$ such that $\int_0^1 \widetilde{c}(x)\,dx = 1$. Alfonsi and Brigo (2005) show that this density actually yields a copula C which they call a *periodic copula*. Note that such a periodic copula can be asymmetric (non-exchangeable) when $\widetilde{c}$ is not an even function. Let us consider two examples in which classes of asymmetric copulas are identifiable in the context of the competing risks model in the special case where the margins are uniform. The construction of these classes is according to Alfonsi and Brigo (2005).

Example 2.8 (Periodic jump copulas) Suppose that X and Y are uniformly distributed on the unit interval. For $\gamma \in (0,1/2)$, let $\widetilde{c}_\gamma$ be the periodic continuation on $\mathbb{R}$ of $\gamma^{-1}\mathbf{1}_{[0,\gamma]}(x)$ (with $0 \le x < 1$), denote by C_γ the corresponding periodic copula, and let $\mathcal{C} := \{C_\gamma \mid \gamma \in (0,1/2)\}$. It is easy to see that

$$C_{x,\gamma}(z) := \frac{\partial}{\partial x} C(x,z)\Big|_{x=z} = \frac{z}{\gamma} \wedge 1.$$

This implies that for two different $\gamma, \gamma' \in (0,1/2)$, we have that $C_{x,\gamma}(\gamma \wedge \gamma') \neq C_{x,\gamma'}(\gamma \wedge \gamma')$, whence the identifiability of $\mathcal{C}$.

There are also classes of asymmetric copulas that are not identifiable. We end this section with such a negative example.

Example 2.9 (Generalised Cuadras-Augé family) This class is defined by

$$C_{\alpha,\beta}(u,v) = \begin{cases} u^{1-\alpha}v & u^\alpha \ge v^\beta \\ uv^{1-\beta} & u^\alpha \le v^\beta. \end{cases}$$

for $0 < \alpha, \beta < 1$ (cf. Nelsen, 2006, p.52ff). The domain of such a copula is divided into two parts by the graph of the function $f_{\alpha,\beta}(x) = x^{\alpha/\beta}$. In the upper left part of the domain, the copula only depends on the parameter β, in the lower right part only on α. Thus, the parameter α is obviously determined by the value of the copula at any single point in the lower right part of the domain, and the parameter β by such a value in the other part of the domain.

Recall that by virtue of (2.2), the copula's values along the curve Ψ defined in (2.3) are identifiable based on the distribution of (Z,Δ). In view of the above discussion this means that a copula's parameters α and β can be identified if the curve Ψ crosses the graph of the function $f_{\alpha,\beta}$ and, more precisely, there are two points on Ψ of which we know that they lie on opposite sides of the graph of $f_{\alpha,\beta}$. Observe that the curve Ψ is the graph of the function

$$\psi : [0,1] \to [0,1] : \quad u \mapsto F_Y(F_X^{-1}(u))$$

and suppose that the marginal distributions F_X and F_Y are such that the expression $\log \psi(u)/\log u$ is nowhere constant as a function of u. The latter condition amounts to assuming that there is no constant $r > 0$ such that $F_X(t) = F_Y(t)^r$ on any interval inside $[0,1]$.

Let $z \in \mathbb{R}$ and $u := u(z) := F_X(z)$. Then, we have that $H(z) = C(u, \psi(u))$ and, by definition of the Cuadras-Augé copulas,

$$H(z) = \begin{cases} u^{1-\alpha}\psi(u) & u^\alpha \geq \psi(u)^\beta \\ u\psi(u)^{1-\beta} & u^\alpha \leq \psi(u)^\beta. \end{cases}$$

Dividing by $u\psi(u)$, taking the logarithm, and finally dividing by $\log u$, we obtain

$$\Gamma(z) := \frac{\log H(z) - \log u - \log \psi(u)}{\log u} = \begin{cases} -\alpha & \psi(u) \leq f_{\alpha,\beta}(u) \\ -\beta \frac{\log \psi(u)}{\log u} & \psi(u) \geq f_{\alpha,\beta}(u). \end{cases}$$

The left hand side of the last equation is identifiable based on the distribution of (Z, Δ) and we may therefore consider it as known. The right hand side is constant for $\psi(u) \leq f_{\alpha,\beta}(u)$ and nowhere constant otherwise by assumption. Suppose that the curve Ψ lies on both sides of the graph of $f_{\alpha,\beta}$. On the one hand, function Γ is then constant on some intervals where it takes the value $-\alpha$. On the other hand, there are intervals where Γ is nowhere constant and $\Gamma(z)\log(u)/\log(\psi(u)) = -\beta$. As $\log(u)/\log(\psi(u))$ is known, both parameters α and β are identifiable in this case. To summarise, we obtain the following sufficient identifiability condition:

A set of parameters $\mathcal{A} \subset (0,1)^2 \setminus \{(t,t) \mid t \in (0,1)\}$ defining a subclass of the generalised Cuadras-Augé family is identifiable if

$$\forall\, (\alpha, \beta) \in \mathcal{A} \quad \exists\, x_1, x_2 \in (0,1) \ : \ \psi(x_1) > f_{\alpha,\beta}(x_1) \ \text{ and } \ \psi(x_2) < f_{\alpha,\beta}(x_2)$$

and if additionally $\log \psi(u)/\log u$ is nowhere constant as a function of u.

If the latter condition is not satisfied, but we know two points on Ψ that lie on opposite sides of the graph of $f_{\alpha,\beta}$ for every $(\alpha, \beta) \in \mathcal{A}$, then the parameters are still identified, which yields the following condition:

$$\exists\, x_1, x_2 \in (0,1) \quad \forall\, (\alpha, \beta) \in \mathcal{A} \ : \ \psi(x_1) > f_{\alpha,\beta}(x_1) \ \text{ and } \ \psi(x_2) < f_{\alpha,\beta}(x_2).$$

3. One unknown margin

In this section, we consider the case where one of the two margins is unknown. Our objective is to show that the unknown margin F_X and the copula C_{XY} can be identified simultaneously under certain conditions. To this end, we need stronger assumptions than in the previous section, where both margins were entirely known. More specifically, we will suppose that the two observed variables

$$\Delta \text{ and } Z \text{ are stochastically independent.} \tag{3.1}$$

As far as the family of copulas is concerned, we use the same conditions as in Corollary 2.6. We show below that in this setting, the copula is identifiable.

It is noteworthy that in the context of classical survival analysis, i.e. assuming independence of the competing risks X and Y, the additional condition (3.1) is equivalent to assuming the so-called Koziol-Green model. In this censoring model, one assumes that the survival function of the event time X is a power of the survival function of the censoring time Y. The Koziol-Green model has been applied successfully in some practical situations. For example, Koziol and Green (1976) and Csörgő and Horváth (1981) consider prostate cancer data, whereas Csörgő (1988) treats the Channing House data from Hyde (1977). Nevertheless, the classical Koziol-Green model has been criticised as unrealistic in many settings, e.g. Csörgő and Faraway (1998) called it "too good to be frequently true". Numerous modifications of the original model have been suggested, e.g. the Generalised Koziol-Green model in which hypothesis (3.1) is weakened by assuming that Δ and Z may be dependent, but according to a known copula. It is not the scope of this paper to contribute to the debate on the usefulness of the Koziol-Green model and its modifications. Our purpose is rather the extension of the identification results of the previous section to the context of dependent competing risks with one unknown margin.

When X and Y are allowed to be dependent, condition (3.1) is weaker than the original "proportional hazards" assumption made by Koziol and Green (1976). Braekers and Veraverbeke (2008) consider this case and develop a strongly consistent estimator of F_X in this context, even when F_Y is unknown. However, they assume the copula C_{XY} to be known. The next result shows that under assumption (3.1), the distribution function F_X and the copula C_{XY} can be identified simultaneously (the latter within certain parametric classes).

Theorem 3.1 *Suppose that Δ and Z are independent and that F_Y is known. Assume further that both F_X and F_Y are continuous and strictly increasing in $(0, \infty)$. As in Corollary 2.6, let $\Phi := \{\varphi_\theta \mid \theta \in \Theta \subset \mathbb{R}\} \subset \Omega$ be a family of bounded generators such that $\theta \mapsto \varphi_\theta(t)/\varphi_\theta(0)$ is strictly monotonic in θ for every $t \in (0, 1)$. Then, the unknown marginal law F_X and the copulas generated by the class Φ are identifiable.*

Proof. For given $z \in (0,1)$, the level curve of level z intersects with the curve $\Psi = (F_X(t), F_Y(t))_{t\in\mathbb{R}}$ for

$$t = F_X^{-1}\Big[\varphi_\theta^{[-1]}\Big\{-\int_{H^{-1}(z)}^1 \varphi_\theta'(H(s))\,dH(s,1)\Big\}\Big]$$

or equivalently at the point

$$\Big(\varphi_\theta^{[-1]}\Big\{-\int_{H^{-1}(z)}^1 \varphi_\theta'(H(s))\,dH(s,1)\Big\}, \varphi_\theta^{[-1]}\Big\{-\int_{H^{-1}(z)}^1 \varphi_\theta'(H(s))\,dH(s,0)\Big\}\Big)$$
$$:= (\zeta_x^\theta(z), \zeta_y^\theta(z)). \tag{3.2}$$

In order to prove (3.2), we need to show that $C(\zeta_x^\theta(z), \zeta_y^\theta(z)) = z$. Since $C(x,y) = \varphi_\theta^{[-1]}(\varphi_\theta(x) + \varphi_\theta(y))$, this means that we need to show that $\varphi_\theta(\zeta_x^\theta(z)) + \varphi_\theta(\zeta_y^\theta(z)) = \varphi_\theta(z)$. Indeed,

$$\begin{aligned}
&\varphi_\theta(\zeta_x^\theta(z)) + \varphi_\theta(\zeta_y^\theta(z)) \\
&= -\int_{H^{-1}(z)}^1 \varphi_\theta'(H(s))\,dH(s,1) - \int_{H^{-1}(z)}^1 \varphi_\theta'(H(s))\,dH(s,0) \\
&= -\int_{H^{-1}(z)}^1 \varphi_\theta'(H(s))\,dH(s) \\
&= -\int_z^1 \varphi_\theta'(u)\,du = \varphi_\theta(z),
\end{aligned}$$

since $\varphi_\theta(1) = 0$. The independence of Δ and Z implies that there is a constant $\alpha \in [0,1]$ such that

$$\begin{aligned}
H(s,0) &= \alpha H(s) \\
H(s,1) &= (1-\alpha)H(s),
\end{aligned}$$

and thus, using (3.2), we can write $\zeta_y^\theta(z) = \varphi_\theta^{[-1]}(\alpha\varphi_\theta(z))$. Following the proof of Corollary 2.6, we obtain that $\zeta_y^\theta(0)$ is strictly monotonic in θ under the assumptions.

Now let $\xi := \sup\{t \in [0,1] \mid C(F_X(t), F_Y(t)) = 0\}$. Note that in view of (2.1), ξ is identifiable based on the distribution of (Z, Δ). Furthermore, we have $\zeta_y^\theta(0) = F_Y(\xi)$, which then implies the identifiability of θ using the monotonicity of $\zeta_y^\theta(0)$ in θ. Having identified θ, we may consider the copula as known. The result then follows applying Theorem 3.1 from Zheng and Klein (1995). □

4. Discussion

We have considered the problem of identifiability in a bivariate competing risks model with dependence between the two event times X and Y. It is well known that without further restrictions on the model, the joint distribution of the event times is not identifiable when only the minimum Z of the two times and the indicator Δ are observed.

First, we showed as an intermediate step that when the marginal distributions of X and Y are known, their copula is identifiable within many popular families of Archimedean copulas, but also in some less known families. However, in many applications at least one of the marginal distributions is the object of interest that has to be estimated.

In the case where only one of the two margins is known, we were still able to establish an identification result, but we needed the additional assumption that Z and Δ are stochastically independent. Although this hypothesis can be tested in practice, it would be desirable to replace it by a weaker assumption. One could imagine that Z and Δ are dependent via some copula $C_{Z\Delta}$, but it is not obvious in which way this copula has to be related to the copula C_{XY} and how identifiability can be shown in this setting.

The problem becomes still harder when both margins are completely unknown. Without any further assumptions on the margins, the class of all Archimedean copulas is too large as to allow for identification (Wang, 2012), and at present, we do not know if identification is possible for certain subfamilies of Archimedean copulas. One could try to tackle this problem assuming the margins to lie in parametric classes.

The generalisation of the results in this paper to the multivariate case should be straightforward, at least for the Archimedean families of copulas. Further research for the development of estimation methods in the models that we have shown to be identifiable.

A. Appendix

Lemma A.1 *If $\{f_\theta\}$ is a family of monotonic bijective functions $f_\theta : [0,1] \to [0,1]$ such that $f_\theta(t)$ is strictly increasing in θ for every $t \in (0,1)$, then $f_\theta^{-1}(t)$ is also strictly increasing in θ for every $t \in (0,1)$.*

Proof. For $\theta' > \theta$ and a given $t \in (0,1)$, let $z := f_{\theta'}^{-1}(t)$. Then, we have $t = f_{\theta'}(z)$, and consequently

$$f_{\theta'}^{-1}(t) > f_\theta^{-1}(t) \iff z > f_\theta^{-1}(f_{\theta'}(z)) \iff f_\theta(z) < f_{\theta'}(z), \tag{A.1}$$

which is true by assumption (see Figure 3 for an illustration of the second inequality). □

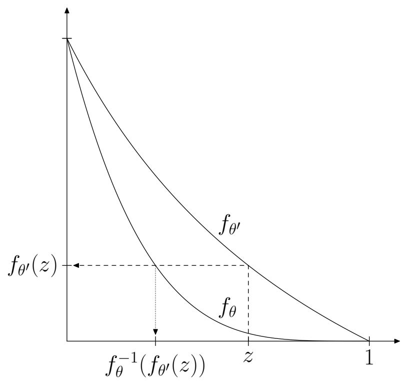

Figure 3: Illustration of inequality (A.1)

Acknowledgments

I. Van Keilegom and M. Schwarz acknowledge financial support from the IAP research network P7/06 of the Belgian Government (Belgian Science Policy) and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650. I. Van Keilegom acknowledges further financial support from the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain".

References

Alfonsi, A. and Brigo, D. (2005). New families of copulas based on periodic functions. *Communications in Statistics, Theory and Methods*, 34(7):1437–1447.

Basu, A. and Ghosh, J. (1978). Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis*, 8:413–429.

Braekers, R. and Veraverbeke, N. (2008). A conditional Koziol-Green model under dependent censoring. *Statistics and Probability Letters*, 78(7):927–937.

Cox, D. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society, Series B*, 21:411–421.

Crowder, M. (1991). On the identifiability crisis in competing risks analysis. *Scandinavian Journal of Statistics*, 18(3):223–233.

Csörgő, S. (1988). Estimation in the proportional hazards model of random censorship. *Statistics*, 19(3):437–463.

Csörgő, S. and Faraway, J. J. (1998). The paradoxical nature of the proportional hazards model of random censorship. *Statistics*, 31(1):67–78.

Csörgő, S. and Horváth, L. (1981). On the Koziol-Green model for random censorship. *Biometrika*, 68:391–401.

Ebrahimi, N., Molefe, D., and Ying, Z. (2003). Identifiability and censored data. *Biometrika*, 90(3):724–727.

Gumbel, E. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55:698–707.

Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika*, 64:22–30.

Klein, J. P. and Moeschberger, M. (1988). Bounds on net survival probabilities for dependent competing risks. *Biometrics*, 44(2):529–538.

Koziol, J. A. and Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika*, 63:465–474.

Marshall, A. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30–44.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics, second edition.

Rivest, L.-P. and Wells, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *Journal of Multivariate Analysis*, 79(1):138–155.

Slud, E. V. and Rubinstein, L. V. (1983). Dependent competing risks and summary survival curves. *Biometrika*, 70:643–649.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences USA*, 72(1):20–22.

Wang, A. (2012). On the nonidentifiability property of archimedean copula models under dependent censoring. *Statistics and Probability Letters*, 82(3):621 – 625.

Wysocki, W. (2012). Constructing archimedean copulas from diagonal sections. *Statistics and Probability Letters*, 82(4):818 – 826.

Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82(1):127–138.