Master's Seminar:
The Copula-Graphic Estimator.
WS 2022/23

Claudio Longo

27.01.2023

This work is based on [4]. Many passages are very close to this source.

Contents

0 Introduction

In this paper, we observe two competing risks A and B and their respective times to occurrence $X \geq 0$ and $Y \geq 0$. Given the observed sample data of X and Y, we want to estimate the marginal survival functions under the assumption that the dependency structure of X and Y can be described by a known copula C.

In this context "competing" means that only event A or B can be observed. If $Y \leq X$, then event B occurred before event A and therefore we can't observe X any more.

In a clinical study X may be the time to death (risk A) and Y the time to withdrawal of the subject from the study (risk B). If we are only interested in X, then Y can be regarded as censoring the event of interest.

Definition 0.1 (censoring, time to censoring)**.**
In the context of survival analysis, censoring describes a missing data problem where the time to an event of interest is not observed. Assuming that we view A as the event of interest, then Y is the time to censoring. If the subject withdraws from the study ($Y \leq X$), then time to death (X) is not observed.

This is called right-censoring - we only observe the lower limit Y for X.

0.1 Probabilistic Setup

To model this, we consider a probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with $A, B \in \mathcal{F}$ and a stochastic process $(I_t)_{t\geq 0}$ adapted to $(\mathcal{F}_t)_{t\geq 0}$. The process I_t can be viewed as an indicator whether an event of interest occurred and using this, we define $X = \inf\{t \geq 0 : I_t \in A\} \sim F$ and $Y = \inf\{t \geq 0 : I_t \in B\} \sim G$ to be the hitting times of A and B.

0.2 Sample data

Since we consider competing risks A and B, we can only observe X or Y. If Y is observed, then X is censored and vice versa. Hence, the observable sample information is

i. time to observation of either A or B, given by $T = \min(X, Y)$ and
ii. information about which event was observed, captured by $\delta = \mathbb{1}_{\{X \leq Y\}}$.

We conclude that the observable sample information is given by (T, δ).

In the competing-risk framework we observe $(T_1, \delta_1), \ldots, (T_n, \delta_n)$. Using this data and assuming that $\mathbb{P}(X = Y) = 0$, we can directly estimate the following quantities

$$k(t) = \mathbb{P}(X > t, Y > t),\ p_1(t) = \mathbb{P}(X \leq t, X < Y) \text{ and } p_2(t) = \mathbb{P}(Y \leq t, Y < X).$$

Note that we have $p_1(t) = \mathbb{P}(T \leq t, \delta = 1)$ and $p_2(t) = \mathbb{P}(T \leq t, \delta = 0)$.

1 A common assumption: Y independent of X

In order to estimate the marginal survival function S_X of X, a common assumption is that the censoring time Y is independent of X.

Remark (under independence F is uniquely determined by the observable data (T, δ))**.** In this case the observable data (T, δ) provides enough information to uniquely determine the marginal survival function of X. In particular Berman showed in [1] that under the assumption of independence of X and Y, we can rewrite the marginal distribution F of X as a function of the observable quantities p_1 and p_2

$$F(t) = 1 - \exp\Big\{ - \int_{-\infty}^{t} \big(1 - [\, p_1(\tau) + p_2(\tau)] \,\big)^{-1} \, dp_1(\tau) \Big\}.$$

When independence is assumed, a standard estimator for the marginal survival function S_X is given by the Kaplan-Meier estimator.

Definition 1.1 (Kaplan-Meier estimator for censored data)**.**
Let x_k be the observed value of X_k if there is no censoring $(T_k = x_k, \delta_k = 1)$ and let y_k be its value in case of censoring $(T_k = y_k, \delta_k = 0)$. Note that y_k is the censoring time of X_k.

Let $t_1, \ldots, t_m$ be the m unique values of the x_k that are observed in the sample data and let s_k be the number of times that the uncensored observation t_k appears in the sample.

Let r_k be the risk set at the k-th observation

$$r_k = |\{j : x_j \geq t_k\}| + |\{j : y_j \geq t_k\}|.$$

The risk set comprises the observations

i. where X is uncensored and observed after time t_k $(Y_k > X_k \geq t_k)$,
ii. that get censored at a time later than t_k $(X_k > Y_k \geq t_k)$.

Intuition:
Just before time t_k there are r_k data points where neither risk A nor B was observed. Those are "at risk", meaning that event A or B could potentially occur at time t_k. Moving on to time t_k we observe s_k occurrences of either risk A or B. The probability of surviving past t_k is therefore given by

$$\frac{r_k - s_k}{r_k} = 1 - \frac{s_k}{r_k}.$$

The Kaplan-Meier estimator for censored data is given by

$$\hat{S}_{KM} = \prod_{k:\, t_k \leq t} \Big(1 - \frac{s_k}{r_k}\Big).$$

We need to be careful with this assumption, since independence may not always be true. For instance, suppose we consider a callable bond with the two competing risks A: "bond defaults" and B: "bond is called". Then independence of X and Y does not hold true.

However, we need to make some assumption about the relation of X and Y in order to identify the marginal distribution of X from the competing risk data we observe.

2 Assumption proposed by the paper: known copula

The paper suggests a different approach that **does not assume independence** of X and Y. Instead we make other assumptions regarding the marginal distributions and the dependency structure. First, we assume that the marginal distributions are strictly increasing and continuous. In this case, it follows from Sklar's Theorem that there exists a **unique** copula C, such that the joint distribution function H of $X \sim F$ and $Y \sim G$ can be represented by

$$H(y_1, y_2) = C(F(y_1), G(y_2)) \quad \forall y_1, y_2 \in \mathbb{R}.$$

Further, **we make the assumption that the copula C of X and Y is known**.

Remark (A reasonable assumption?)**.**
This assumption about the dependency structure of the two competing risks is untestable, but so is the independence assumption.

Definition 2.1 (copula)**.**
A function $C : [0,1]^d \to [0,1]$ is called copula, if there exists a random vector $(Y_1, \ldots, Y_d)$ such that the following properties hold

i. $Y_j \sim U([0,1])$, $\forall j = 1, \ldots, d$ (uniform margins)
ii. C is the joint distribution of $(Y_1, \ldots, Y_d)$
$$C(y_1, \ldots, y_d) = \mathbb{P}(Y_1 \leq y_1, \ldots, Y_d \leq y_d),\ \forall y_1, \ldots, y_d \in [0,1].$$

Note that C itself is a distribution function. We define the probability measure μ_C by

$$\mu_C(E) = \int_{[0,1]^d} \mathbb{1}_E \, dC, \text{ for } E \subseteq [0,1]^d.$$

From now on, we restrict ourselves to bivariate copulas ($d = 2$).

2.1 How copulas and the dependency structure relate

The copula captures the dependency structure between the margins X and Y. In case of independence, the copula is given by $C_I(y_1, y_2) = y_1 y_2$. Extreme linear dependency can be described by the comonotonicity copula $M_2(y_1, y_2) = \min\{y_1, y_2\}$ and the countermonotonicity copula $W_2(y_1, y_2) = \max\{y_1 + y_2 - 1, 0\}$. We have that

i. M_2 encodes the most positive dependence: $(Y_1, Y_2) \sim M_2 \iff \exists a > 0 : Y_1 \stackrel{d}{=} a\, Y_2$
ii. W_2 captures the most negative dependence: $(\tilde{Y}_1, \tilde{Y}_2) \sim W_2 \iff \exists \tilde{a} < 0 : \tilde{Y}_1 \stackrel{d}{=} \tilde{a}\, \tilde{Y}_2$
iii. Fréchet-Hoeffding bounds: $W_2(y_1, y_2) \leq C(y_1, y_2) \leq M_2(y_1, y_2)$, $\forall y_1, y_2 \in [0,1]$.

Further we can rewrite the most common non-parametric dependency measures such as Kendall's τ or Spearman's ρ as functions of the copula of X and Y. In particular non-parametric dependency measures are normed distances of the copula of X and Y from the independence copula.

3 Identifiability

3.1 Theorem 3.1

Theorem 3.1 (Theorem 3.1)**.**
Suppose the marginal distribution functions of (X, Y) *are continuous and strictly increasing in* $(0, \infty)$*. Suppose the copula* C *of* (X, Y)*, is known, and* $\mu_C(E) > 0$ *for any open set* $E \in [0,1] \times [0,1]$*. Then* F *and* G*, the marginal distribution functions of* X *and* Y*, are uniquely determined by* $\{k(t), p_1(t), p_2(t) \mid t > 0\}$*.*

Remark (under assumed copula C, the observable data (T, δ) uniquely determines F)**.**
This Theorem states that if the copula is assumed to be known, then the marginals F and G are uniquely determined by $\{k(t), p_1(t), p_2(t) \mid t > 0\}$. These quantities can be directly estimated from the observable data.

Notation.
Before we give a proof of Theorem 3.1, we introduce some notation and derive an alternative representation of p_1. For $t > 0$,

$$p_1(t) = \mathbb{P}(X \leq t, X < Y) = \int_{\mathbb{R}^2} \mathbb{1}_{C_t} \, dHt,$$

where $C_t = \{(x, y) \mid x < y, 0 < x < t\}$. Further,

$$\begin{aligned} p_1(t) &= \int_{\mathbb{R}^2} \mathbb{1}_{C_t}(x, y) \, dH(x, y) \\ &= \int_{[0,1]^2} \mathbb{1}_{C_t}(F^{-1}(u), G^{-1}(v) \, dH(F^{-1}(u), G^{-1}(v)) \\ &= \int_{[0,1]^2} \mathbb{1}_{C_t}(F^{-1}(u), G^{-1}(v)) \, dC(u, v). \end{aligned}$$

Let $(x, y) \in [0,1]^2$ such that $(F^{-1}(x), G^{-1}(y)) \in C_t$. Then

$$\begin{aligned} &0 < F^{-1}(x) < t \iff F(0) = 0 < x < F(t) \text{ and} \\ &F^{-1}(x) < G^{-1}(y) \iff G(F^{-1}(x)) < y < 1. \end{aligned}$$

It follows that

$$(F^{-1}(x), G^{-1}(y)) \in C_t \iff (x, y) \in \{(u, v) \mid 0 < u < F(t), GF^{-1}(u) < y < 1\} =: B_{(F,G)_t}.$$

We can rewrite p_1 as

$$p_1(t) = \int_{[0,1]^2} \mathbb{1}_{C_t}(F^{-1}(u), G^{-1}(v)) \, dC(u, v) = \mu_C(B_{(F,G)_t}).$$

The proof of Theorem 3.1 uses the following Lemma.

3.2 Lemma A1

Lemma 3.2 (Lemma A1)**.**
Suppose that the marginal distribution functions of (X, Y) *are continuous and strictly increasing in* $(0, \infty)$*, and* $\mu_c(E) > 0$ *for any open set* $E \in (0,1) \times (0,1)$*. Also suppose that* (F, G) *and* (F_1, G_1) *have the same set of* $\{k(t), p_1(t), p_2(t) \mid t > 0\}$*. If there is a point* $0 < x_0 < 1$ *such that* $G_1 F_1^{-1}(x_0) = G F^{-1}(x_0)$ *then* $F_1^{-1}(x_0) = F^{-1}(x_0)$*.*

Proof of Lemma A1.
By assumption, we find $x_0 \in (0,1)$ such that $G_1(F_1^{-1}(x_0)) = G(F^{-1}(x_0))$. Further, since we assumed that $k(t)$ is the same for both pairs of margins, it follows that

$$k(t) = \mu_C(A_t) = \mu_C(A_t^1) \quad \forall t > 0.$$

In particular for $t_0 = F^{-1}(x_0)$, we have $\mu_C(A_{t_0}) = \mu_C(A_{t_0}^1)$.

Suppose that $F_1^{-1}(x_0) > F^{-1}(x_0) = t_0$. We want to show that in this case we have $A_{t_0} \subsetneq A_{t_0}^1$ and hence $\mu_C(A_{t_0}) < \mu_C(A_{t_0}^1)$, contradicting our assumption regarding $k(t)$.

Ad A_{t_0} and $A_{t_0}^1$:
A_{t_0} is a rectangle with vertices $(x_0, G(F^{-1}(x_0))), (1, G(F^{-1}(x_0))), (1,1), (x_0, 1)$.

$A_{t_0}^1$ is a rectangle with vertices $(F_1(t_0), G_1(t_0)), (1, G_1(t_0)), (1,1), (F_1(t_0), 1)$.

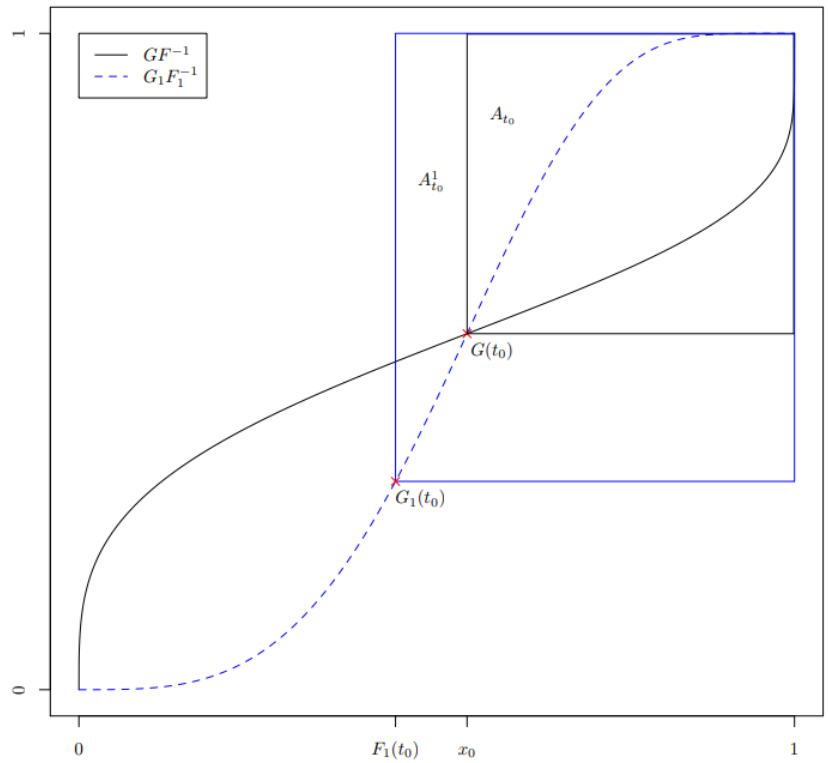


Figure 1: The case of $F_1^{-1}(x_0) > t_0$.

Now $A_{t_0} \subsetneq A^1_{t_0}$ holds, if we can show that $F_1(t_0) < x_0$ and $G_1(t_0) < G(t_0)$. First note that the first condition holds by assumption

$$F_1^{-1}(x_0) > t_0 \Longleftrightarrow x_0 > F_1(t_0).$$

So we only need to show $G_1(t_0) < G(t_0)$. Here we can use that $G_1(F_1^{-1}(t_0)) = G(F^{-1}(t_0))$

$$\begin{aligned} G_1(t_0) < G(t_0) &\Longleftrightarrow G_1(F^{-1}(x_0)) < G(F^{-1}(x_0)) = G_1(F_1^{-1}(x_0)) \\ &\Longleftrightarrow G_1(F^{-1}(x_0)) < G_1(F_1^{-1}(x_0)) \\ &\Longleftrightarrow F^{-1}(x_0) < F_1^{-1}(x_0) \\ &\Longleftrightarrow F_1(F^{-1}(x_0)) < F_1(F_1^{-1}(x_0)) \\ &\Longleftrightarrow F_1(t_0) < x_0. \end{aligned}$$

We conclude the contradiction $A_{t_0} \subsetneq A^1_{t_0}$.

This shows that $F_1^{-1}(x_0) \leq F^{-1}(x_0)$ must hold. Further, by interchanging the roles of F and F_1, we see that $F_1(x_0) = F(x_0)$ which concludes the proof.

□

3.3 Proof of Theorem 3.1

Proof of Theorem 3.1.
Proof by contradiction. Suppose there exists $t_0 > 0$ such that $x_1 = F_1(t_0) < F(t_0)$.

The proof consists of four main steps.

1. First, we want to show that

$$G_1(t_0) > G(t_0). \tag{1}$$

Suppose that $G_1(t_0) \leq G(t_0)$. Then we have for the vertices $(F_1(t_0), G_1(t_0)) < (F(t_0), G(t_0))$ and hence $A_{t_0} \subsetneq A^1_{t_0}$. It follows that

$$\mu_C(A_{t_0}) < \mu_C(A^1_{t_0}),$$

since $\mu_C(E) > 0$ for all open sets $E \in [0,1] \times [0,1]$.

This contradicts $\mu_C(A^1_{t_0}) = k(t) = \mu_C(A_{t_0})$ and we conclude $G_1(t_0) > G(t_0)$.

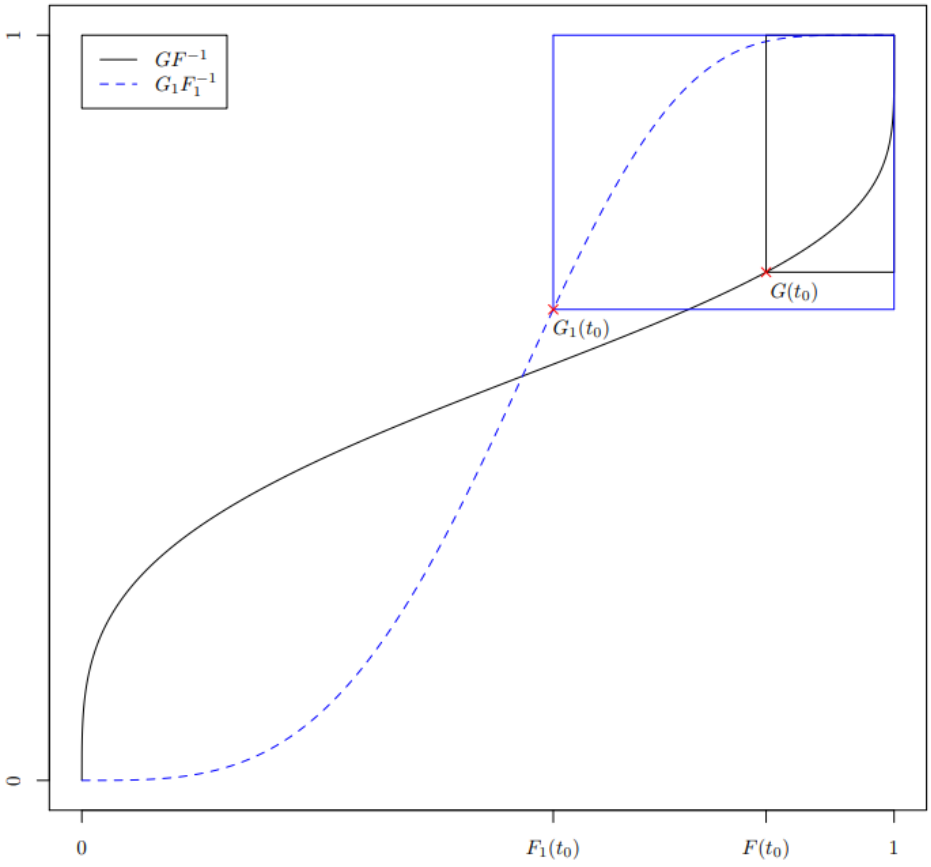


Figure 2: Situation for $G_1(t_0) > G(t_0)$.

We want to illustrate the positions of $F_1(t_0), F(t_0)$ and $G_1(t_0), G(t_0)$ in this setting.

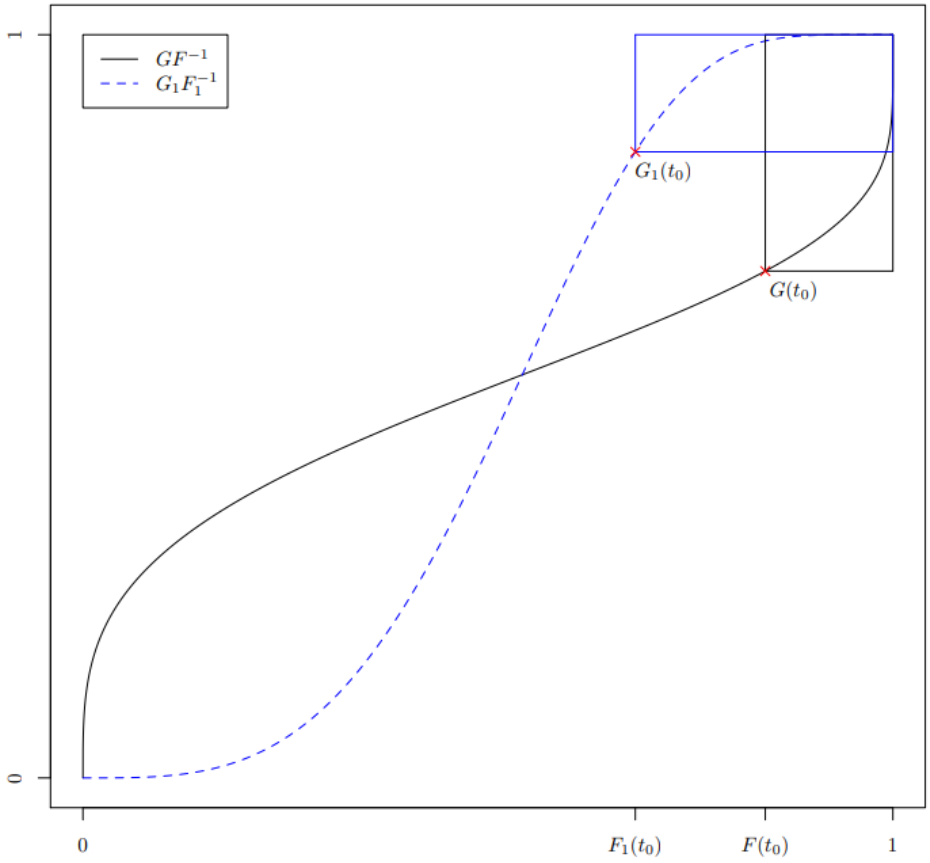


Figure 3: Positions of $F_1(t_0), F(t_0)$ and $G_1(t_0), G(t_0)$.

2. Next, we argue that

$$G(F^{-1}(x_1)) < G_1(F_1^{-1}(x_1)). \tag{2}$$

It is easy to see that $G(F^{-1}(x_1)) < G(t_0) < G_1(F_1^{-1}(x_1))$ must hold.

Recall that F and G are strictly increasing. Hence, we have that

$$\underbrace{F_1(t_0)}_{=x_1} < F(t_0) \iff F^{-1}(x_1) < t_0 \iff G(F^{-1}(x_1)) < G(t_0).$$

It follows that

$$G(F^{-1}(x_1)) < G(t_0) \overset{(1)}{<} G_1(t_0) = G_1(F_1^{-1}(x_1)).$$

3. show that $G(F^{-1}(x)) \leq G_1(F_1^{-1}(x))$ can't hold for all $x \in (0, x_1)$

Suppose $G(F^{-1}(x)) \leq G_1(F_1^{-1}(x))$ holds for all $x \in (0, x_1)$.

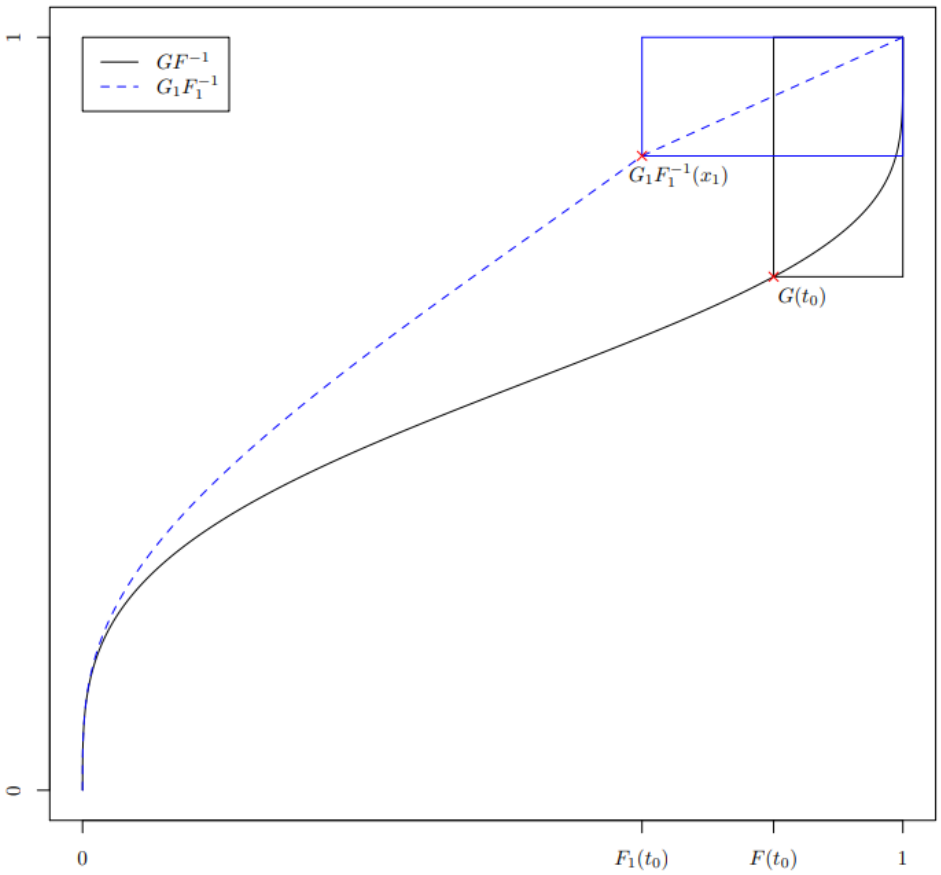


Figure 4: Situation when $G(F^{-1}(x)) \leq G_1(F_1^{-1}(x))$ for all $x \in (0, x_1)$

Idea: show that $B_{(F_1,G_1)_{t_0}} \subsetneq B_{(F,G)_{t_0}}$ and hence $\mu_C(B_{(F_1,G_1)_{t_0}}) < \mu_C(B_{(F,G)_{t_0}})$.

$$B_{(F_1,G_1)_{t_0}} = \{(x,y) \mid 0 \leq x < \overbrace{F_1(t_0)}^{x_1=}, G_1F_1^{-1}(x) < y < 1\}$$

Recall that for all $x \in (0, x_1) : GF^{-1}(x) \leq G_1F_1^{-1}(x)$, which yields the following inclusion

$$\subset \{(x,y) \mid 0 \leq x < x_1, GF^{-1}(x) < y < 1\}$$

By assumption, we have that $x_1 < F(t_0)$ and hence

$$\begin{aligned} &\subsetneq \{(x,y) \mid 0 \leq x < F(t_0), GF^{-1}(x) < y < 1\} \\ &= B_{(F,G)_{t_0}}. \end{aligned}$$

We find that

$$\begin{aligned} B_{(F,G)_{t_0}} \setminus B_{(F_1,G_1)_{t_0}} = &\{(x,y) \mid 0 \leq x < F_1(t_0), GF^{-1}(x) < y \leq G_1F_1^{-1}(x)\} \\ &\cup \{(x,y) \mid F_1(t_0) \leq x \leq F(t_0), GF^{-1}(x) < y < 1\} \end{aligned}$$

Noting that $\mu_C(B_{(F,G)_{t_0}} \setminus B_{(F_1,G_1)_{t_0}}) > 0$, we come up with the following contradiction

$$p_1(t_0) = \mu_C(B_{(F_1,G_1)_{t_0}}) < \mu_C(B_{(F,G)_{t_0}}) = p_1(t_0).$$

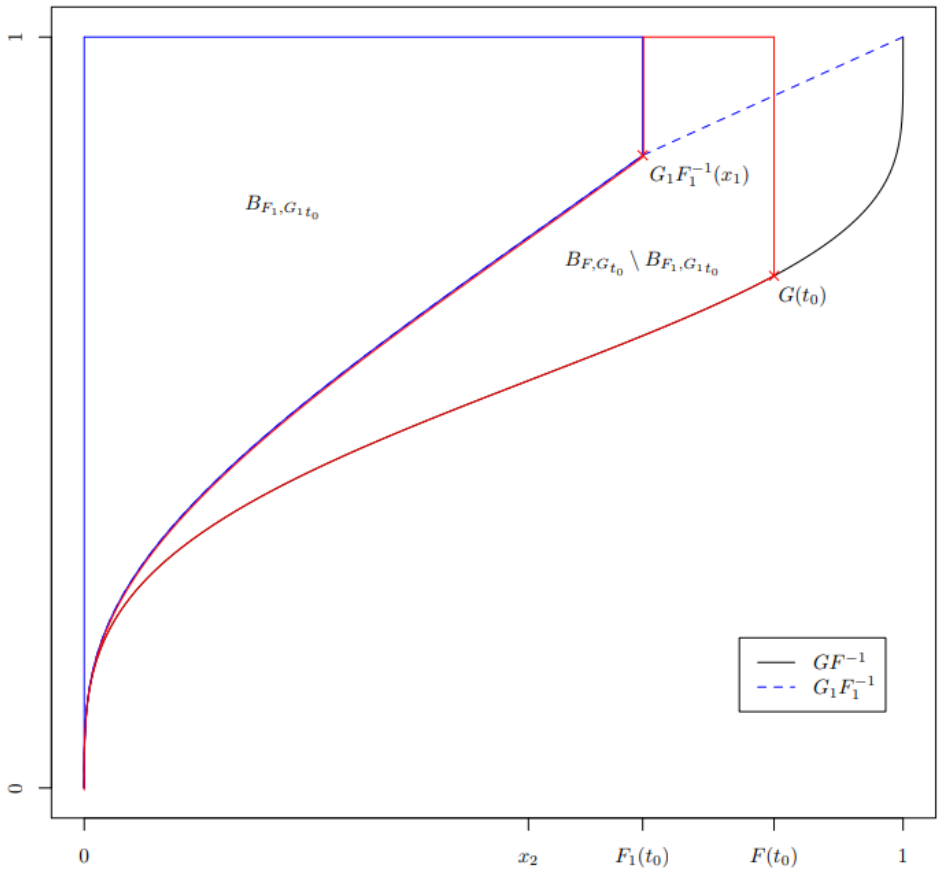


Figure 5: Illustration of the sets $B_{(F_1,G_1)_{t_0}}$ and $B_{(F,G)_{t_0}} \setminus B_{(F_1,G_1)_{t_0}}$

It follows that there exists at least one crossing point $\tilde{x} \in (0, x_1)$ of GF^{-1} and $G_1F_1^{-1}$.

4. the contradiction

Let $x_2 \in (0, x_1)$ be the last crossing point before x_1, such that

$$GF^{-1}(x_2) = G_1F_1^{-1}(x_2) \text{ and } GF^{-1}(x) \leq G_1F_1^{-1}(x) \quad \forall x \in (x_2, x_1).$$

Now, by Lemma A1 we have that $F^{-1}(x_2) = F_1^{-1}(x_2) =: t_1$. Further, we have

$$\begin{aligned}
\mathbb{P}(t_1 < X \leq t_0, X \leq Y) &= p_1(t_0) - p_1(t_1) \\
&= \mu_C(B_{(F,G)_{t_0}} \setminus B_{(F,G)_{t_1}}) \\
&= \mu_C(B_{(F_1,G_1)_{t_0}} \setminus B_{(F_1,G_1)_{t_1}}).
\end{aligned}$$

We can show that $B_{(F_1,G_1)_{t_0}} \setminus B_{(F_1,G_1)_{t_1}} \subsetneq B_{(F,G)_{t_0}} \setminus B_{(F,G)_{t_1}}$ holds.

$$B_{(F_1,G_1)_{t_0}} \setminus B_{(F_1,G_1)_{t_1}} = \{(x,y) \mid x_2 \leq x < x_1, G_1F_1^{-1}(x) < y < 1\}$$

Recall that for all $x \in (x_2, x_1) : GF^{-1}(x) \leq G_1F_1^{-1}(x)$, which yields the following inclusion

$$\subset \{(x,y) \mid x_2 \leq x < x_1, GF^{-1}(x) < y < 1\}$$

By assumption, we have that $x_1 = F_1(t_0) < F(t_0)$ and hence

$$\begin{aligned} &\subsetneq \{(x,y) \mid x_2 \leq x < F(t_0), GF^{-1}(x) < y < 1\} \\ &= B_{(F,G)_{t_0}} \setminus B_{(F,G)_{t_1}}. \end{aligned}$$

We conclude the proof with the following contradiction

$$p_1(t_0) - p_1(t_1) = \mu_C(B_{(F_1,G_1)_{t_0}} \setminus B_{(F_1,G_1)_{t_1}}) < \mu_C(B_{(F,G)_{t_0}} \setminus B_{(F,G)_{t_1}}) = p_1(t_0) - p_1(t_1).$$

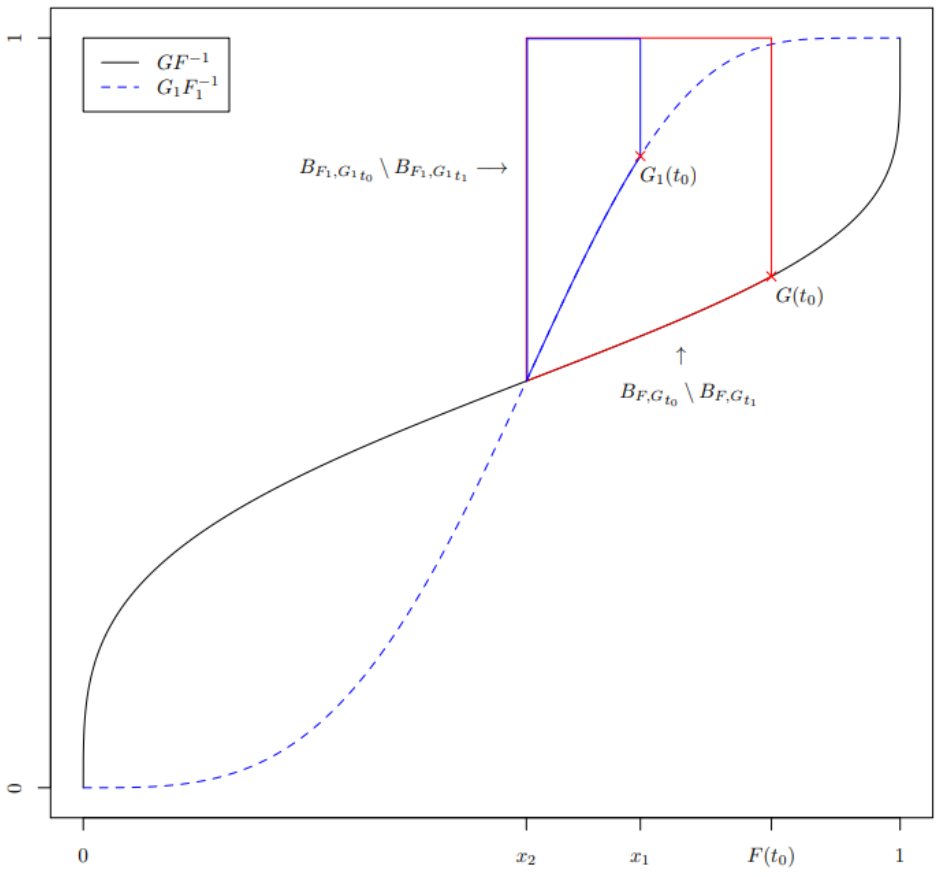


Figure 6: Illustration of the final contradiction

Conclusion.
Thus there is no t_0 such that $F_1(t_0) < F(t_0)$. By exchanging the roles of F_1 and F, we see that $F_1 = F$ must hold. The proof of $G_1 = G$ works the same way.

□

3.4 Corollary 3.2

Remark.
Often it is easier to verify that $u(x, y) > 0$ holds for any $(x, y) \in (0, 1) \times (0, 1)$, than to prove that $\mu_C(E) > 0$ holds for any open set E [3]. This motivates the following Corollary.

Corollary 3.2.1 (Corollary 3.2)**.**
Suppose the marginal distribution functions of (X, Y) are continuous and strictly increasing in $(0, \infty)$. Let $u(x, y)$ be the density function of C.
If $u(x, y) > 0$ for any $(x, y) \in (0, 1) \times (0, 1)$. Then the result of Theorem 3.1 holds.

The proof of Corollary 3.2 is given by the following Lemma.

Lemma 3.3.
If the copula C is absolutely continuous with density function

$$u(x, y) = \frac{\partial^2 C(x, y)}{\partial x \partial y},$$

and $u(x, y) > 0$ for any $(x, y) \in (0, 1) \times (0, 1)$. Then $\mu_C(E) > 0$, for any open set E.

3.5 Corollary 3.3

Remark (marginal distributions with compact support)**.**
Sometimes the condition that F and G are strictly increasing on $(0, \infty)$ is not satisfied. For example when talking about remaining life time distributions, there exists a maximum age that can be achieved. Generally speaking there may exist a time $t_0 > 0$ such that $\mathbb{P}(X > t_0) = 0$ and hence $F(t) = 1 \, \forall t \geq t_0$.

To deal with this situation, we introduce the following corollary.

Corollary 3.3.1 (Corollary 3.3)**.**
Suppose the marginal distribution functions of (X, Y) are continuous and that there are times t_1 and t_2 such that $F(t_1) = 1$ and $G(t_2) = 1$. Further, suppose F and G are strictly increasing in $(0, t_1)$ and $(0, t_2)$ respectively. Suppose the copula C of (X, Y), is known, and $\mu_C(E) > 0$ for any open set $E \in [0, 1] \times [0, 1]$. Then F and G, the marginal distribution functions of X and Y, are uniquely determined on $(0, \min(t_1, t_2))$ by $\{k(t), p_1(t), p_2(t) \mid t > 0\}$.

Remark.
Note that F and G are only uniquely determined up to $\tilde{t} = \min(t_1, t_2)$. This is because after $\tilde{t}$ no variables are observed due to the competing risk setup.

Proof of Corollary 3.3.
The proof of Corollary 3.3 follows the same steps as the proof of Theorem 3.1 with one slight different being that t_0 is chosen from the interval $(0, \min(t_1, t_2))$.

□

4 The Copula-Graphic Estimator

From the proof of Theorem 3.1, we see that F and G are uniquely determined by

$$\mu_C(A_t) = \mathbb{P}(X > t, Y > t) = k(t),$$
$$\mu_C(B_t) = \mathbb{P}(X \leq t, X < Y) = p_1(t).$$

Hence, we find estimators $\hat{F}$ and $\hat{G}$ of F and G by iteratively solving the following system of non-linear equations:

$$\mu_C(\hat{A}_{t_i}) - \hat{P}(X > t_i, Y > t_i) = 0,$$
$$\mu_C(\hat{B}_{t_i}) - \hat{P}(X \leq t_i, X < Y) = 0.$$

We define $\hat{F}$ and $\hat{G}$ to be the Copula-Graphic estimators.

Ad $\hat{P}(X > t_i, Y > t_i)$ and $\hat{P}(X \leq t_i, X < Y)$:
For simplicity, we use the empirical estimates.

$$\hat{P}(X > t_i, Y > t_i) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j > t_i\}} \text{ and } \hat{P}(X \leq t_i, X < Y) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j \leq t_i,\, \delta_j = 1\}}.$$

Ad $\hat{A}_{t_i}$ and $\hat{B}_{t_i}$:

$$\hat{A}_{t_i} = \{(x, y) \mid \hat{F}(t_i) < x \leq 1, \hat{G}(t_i) < y \leq 1\},$$
$$\hat{B}_{t_i} = \{(x, y) \mid 0 \leq x \leq \hat{F}(t_i), \hat{G}\hat{F}^{-1}(t_i) < y \leq 1\}.$$

For more details about the implementation, we refer to chapter six.

4.1 Theorem 4.1

Theorem 4.1 (Theorem 4.1)**.**
Suppose that two marginal distribution functions F, G, are continuous and strictly increasing on $(0, \infty)$, and the assumed copula has density function $u(x, y) > 0$ on $[0, 1] \times [0, 1]$.

Then $\hat{F}_n$ and $\hat{G}_n$ are strongly consistent for F and G. That is with probability 1 as $n \to \infty$, $\hat{F}_n(t) \to F(t)$ and $\hat{G}_n(t) \to G(t)$ for all $t \in [0, \infty)$.

Remark (Probabilistic Setup)**.**
We observe the sample $\{(T_1(\omega), \delta_1(\omega)), (T_2(\omega), \delta_2(\omega)), \ldots, (T_m(\omega), \delta_m(\omega)), \ldots\}$.

Let $E_n = \{t_1^n, \ldots, t_{m_n}^n\}$ be a sequence of subdivisions of $[0, \infty)$ such that

$$|E_n| = \sup_{1 \leq k \leq m_n} |t_k^n - t_{k-1}^n| \to 0.$$

Then $E = \bigcup_n E_n$ is a dense subset of $[0, \infty)$. For each n, we have the data

$$\{(T_1(\omega), \delta_1(\omega)), (T_2(\omega), \delta_2(\omega)), \ldots, (T_{m_n}(\omega), \delta_{m_n}(\omega))\},$$

where m_n is the total number of deaths and censoring. We suppose that $m_n \overset{n \to \infty}{\longrightarrow} \infty$.

4.2 Proof of Theorem 4.1

Proof of Theorem 4.1.
The first step will be to show that for any $t_k \in E$,

$$\hat{P}_n(T > t_k) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j > t_k\}} \xrightarrow{a.s.} \mathbb{P}(T > t_k) = \mathbb{P}(X > t_k, Y > t_k) \text{ as } n \to \infty.$$

This result can be obtained from the SLLN. The assumptions of the SLLN hold, since

- By assumption the T_j are i.i.d. (observed sample data) and hence the corresponding indicator functions are independent.
- For any j and $t \in [0, \infty)$, we have $\mathbb{E}[\mathbb{1}_{\{T_j > t\}}] = \mathbb{P}(T > t)$.
- For any j, we have $\text{Var}(\mathbb{1}_{\{T_j > t\}}) < \infty$.

By the SLLN, it follows that

$$\hat{P}_n(T > t_k) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j > t_k\}} \xrightarrow{a.s.} \mathbb{E}[\mathbb{1}_{\{T > t_k\}}] = \mathbb{P}(T > t_k).$$

For t_k, we define $\Omega_k = \{\omega \in \Omega : \hat{P}_n(T > t_k)(\omega) \longrightarrow \mathbb{P}(T > t_k)\}$ with $\mathbb{P}(\Omega_k) = 1$. Further, let

$$\hat{\Omega} = \bigcap_k \Omega_k.$$

$\hat{\Omega}$ is a countable intersection of almost sure events and therefore itself an almost sure event. Now, for any $\omega \in \hat{\Omega}$ and any $t_k \in E$, we have

$$\hat{P}_n(T > t_k)(\omega) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j(\omega) > t_k\}} \longrightarrow \mathbb{P}(T > t_k) \text{ as } n \to \infty.$$

For the following, we fix $\omega \in \hat{\Omega}$ and for convenience, we write $\hat{F}_n(\omega)$ as $\hat{F}_n$, $\hat{G}_n(\omega)$ as $\hat{G}_n$.

The proof of the theorem, consists of the following two steps.

- First, we show that for any sub-sequence $\{n_i \mid i \in \mathbb{N}_{\geq 1}\}$, there is a sub-sub-sequence $\{n_{ij} \mid j \in \mathbb{N}_{\geq 1}\}$ such that $\hat{F}_{n_{ij}} \longrightarrow F_1$ and $\hat{G}_{n_{ij}} \longrightarrow G_1$.
- Second we prove that F_1 must be F and G_1 must be G.

These two steps together complete our proof, since for any t, any sub-sequence

$$\{\hat{F}_{n_1}(t), \hat{F}_{n_2}(t), \ldots \hat{F}_{n_k}(t), \ldots\}$$

must contain a sub-sub-sequence which converges to $F(t)$. For the $\hat{G}_n$ it is the same.

Remark.
At this point we already know that the whole sequence $\hat{F}_n(t)(\omega)$ is convergent. Hence, if there exists a sub-sub-sequence $\{n_{ij} \mid j \in \mathbb{N}_{\geq 1}\}$ such that $\hat{F}_{n_{ij}}(t)(\omega)$ converges to $F(t)$, then we can conclude that the whole sequence converges to $F(t)$.

Ad 1^{st} step.
Suppose $\{n_i \,|\, i \in \mathbb{N}_{\geq 1}\}$ is a sub-sequence of $\{n \,|\, n \in \mathbb{N}_{\geq 1}\}$. Recall that $\hat{F}_n$, $\hat{G}_n$ and $\hat{G}_n \hat{F}_n^{-1}$ are (empirical) distribution functions. By Helly's selection theorem, there exists a sub-sub-sequence $\{n_{ij} \,|\, j \in \mathbb{N}_{\geq 1}\}$ of $\{n_i \,|\, i \in \mathbb{N}_{\geq 1}\}$ and a function F_1 which is non-decreasing and right-continuous such that

$$\hat{F}_{n_{ij}} \longrightarrow F_1, \quad \forall \text{ continuity points of } F_1.$$

Using the same theorem two more times, we find sub-sub-sequences

$$\{n_{ijkl} \,|\, l \in \mathbb{N}_{\geq 1}\} \subset \{n_{ijk} \,|\, k \in \mathbb{N}_{\geq 1}\} \subset \{n_{ij} \,|\, j \in \mathbb{N}_{\geq 1}\},$$

and non-decreasing right-continuous functions G_1 and V_1 such that

$$\hat{G}_{n_{ijk}} \longrightarrow G_1 \text{ and } \hat{G}_{n_{ijkl}} \hat{F}^{-1}_{n_{ijkl}} \longrightarrow V_1, \quad \forall \text{ continuity points of } G_1 \text{ and } V_1.$$

Rename the sequence $\{n_{ijkl} \,|\, l \in \mathbb{N}_{\geq 1}\}$ by $\{n_{ij} \,|\, j \in \mathbb{N}_{\geq 1}\}$. Then for $\{n_{ij} \,|\, j \in \mathbb{N}_{\geq 1}\}$, we have for any continuity point t of F_1, G_1 and V_1,

$$\hat{F}_{n_{ij}}(t) \longrightarrow F_1(t), \hat{G}_{n_{ij}}(t) \longrightarrow G_1(t) \text{ and } \hat{G}_{n_{ij}} \hat{F}^{-1}_{n_{ij}}(t) \longrightarrow V_1(t).$$

Ad 2^{nd} step.

$B_t = \{(x,y) | 0 \leq x \leq F_1(t), V_1(x) \leq y \leq 1\}$ and $A_t = \{(x,y) | F_1(t) \leq x \leq 1, G_1(t) \leq y \leq 1\}$.
Let t be a point on which F_1, G_1 and V_1 are continuous. Then for any point (x,y) in the interior of B_t, we have that

$$\mathbb{1}_{\hat{B}_t}(x,y) \longrightarrow \mathbb{1}_{B_t}(x,y) \text{ as } n_{ij} \to \infty. \tag{3}$$

Proof of (3).
Fix (x,y) in the interior of B_t. Then there exists $N_1 \in \mathbb{N}_{\geq 1}$ such that

$$|\hat{F}_{n_{ij}}(t) - F_1(t)| < |x - F_1(t)| \quad \forall\, n_{ij} \geq N_1$$

and hence $0 \leq x \leq \hat{F}_{n_{ij}}(t)$ for all $n_{ij} \geq N_1$. Similarly one can find $N_2 \in \mathbb{N}_{\geq 1}$ such that $\hat{G}_{n_{ij}} \hat{F}^{-1}_{n_{ij}}(t) \leq y \leq 1$ for any $n_{ij} \geq N_2$. Thus, we find $M = \max(N_1, N_2)$ with $(x,y) \in \hat{B}_t$ for all $n_{ij} \geq M$. This shows $\mathbb{1}_{\hat{B}_t}(x,y) \longrightarrow \mathbb{1}_{B_t}(x,y)$ as $n_{ij} \to \infty$.

□

For $t_k \in E$, by dominated convergence when $n_{ij} \to \infty$,

$$\begin{aligned}\int_{\hat{B}_{t_k}} d\mu_C &= \int \mathbb{1}_{\hat{B}_{t_k}}(x,y)\, d\mu_C = \int \mathbb{1}_{\hat{B}_{t_k}}(x,y)\, u(x,y)\, dxdy \\ &\longrightarrow \int \mathbb{1}_{B_{t_k}}(x,y)\, u(x,y)\, dxdy = \int_{B_{t_k}} d\mu_C.\end{aligned}$$

Now, we have that

$$\begin{aligned}&\int_{\hat{B}_{t_k}} d\mu_C \longrightarrow \int_{B_{t_k}} d\mu_C \quad \text{and} \\ &\int_{\hat{B}_{t_k}} d\mu_C = \hat{\mathbb{P}}(X \leq t_k, X < Y) \longrightarrow \mathbb{P}(X \leq t_k, X < Y).\end{aligned}$$

By the uniqueness of limits, we find that

$$\int_{B_{t_k}} d\mu_C = \underbrace{\mathbb{P}(X \leq t_k, X < Y)}_{p_1(t_k)}. \tag{4}$$

Similarly, one can argue that

$$\int_{A_{t_k}} d\mu_C = \underbrace{\mathbb{P}(X > t_k, Y > t_k)}_{k(t_k)}. \tag{5}$$

Further, since E is dense in $[0, \infty)$, (4) and (5) must hold for all t in $[0, \infty)$.

From the proof of Theorem 3.1, we see that the two relationships in (4) and (5) uniquely determine F and G. It follows that $F_1 = F$ and $G_1 = G$.

□

4.3 Theorem 4.2

Theorem 4.2 (Theorem 4.2)**.**
The copula-graphic estimator is a maximum likelihood estimator.

Proof.
This follows from Theorem 1 of [2]. They show that any estimator which gives mass l/n to all sets $E_i = \{(x_j, y_i) : x_j > y_i\} \cup \{(\max(T_i), y_i)\}$, or $\{(x_i, y_j) : x_i < y_j\} \cup \{(x_i, \max(T_i))\}$, where x_i and y_j are the observed death and censoring times, is a maximum likelihood estimator of the joint measure of the death and censoring distribution on $\mathbb{R}^+ \times \mathbb{R}^+$.

The construction of the Copula-Graphic estimator preserves this property.

□

Theorem 4.3 (Theorem 4.3)**.**
For the independence copula $C(x, y) = xy$, when $t < t_n$, the largest observed time, the copula-graphic estimates of the marginal survival functions are exactly the Kaplan-Meier estimates.

Proof.
Notation.
We define n_i: # at risk just before time t_i ("risk set"), d_i: # of events of X at time t_i, and e_i : # of events of Y at time t_i. Note that we have $n_{i+1} = n_i - d_i - e_i$.

The Kaplan-Meier estimates $\hat{S}_{KM}(t_k)$ of X and $\hat{Z}_{KM}(t_k)$ of Y are given by

$$\hat{S}_{KM}(t_k) = \prod_{i=1}^{k}\Big(\frac{n_i - d_i}{n_i}\Big) = \prod_{i=1}^{k}\Big(1 - \frac{d_i}{n_i}\Big)$$
$$\hat{Z}_{KM}(t_k) = \prod_{i=1}^{k}\Big(\frac{n_i - e_i}{n_i}\Big) = \prod_{i=1}^{k}\Big(1 - \frac{e_i}{n_i}\Big).$$

Induction over k:
$i = k = 1$, w.l.o.g. assume $\delta_1 = 1$:
Using $n_1 = n$, $\mu_C(\hat{A}_{t_i}) = 1 - \hat{F}(t_i) - \hat{G}(t_{i-1}) + C(\hat{F}(t_i), \hat{G}(t_{i-1})) = k(t_i)$[1] and the initial condition $\hat{F}(t_0) = 0 = \hat{G}(t_0)$, we find

$$1 - \hat{F}(t_1) = 1 - \frac{d_1}{n_1} \iff \hat{S}_X(t_i) = 1 - \frac{d_1}{n} = \hat{S}_{KM}(t_1).$$

Induction step $k - 1 \longrightarrow k$:
$i = k$, w.l.o.g. assume $\delta_k = 1$:
Since $\delta_k = 1$, it follows that $\hat{G}(t_{k-1}) = \hat{G}(t_k)$. Further, $e_k = 0$ and hence

$$\hat{Z}_{KM}(t_k) = \hat{Z}_{KM}(t_{k-1}) \overset{A(k)}{=} 1 - \hat{G}(t_{k-1}) = 1 - \hat{G}(t_k) = \hat{Z}(t_k).$$

Next, we find the copula graphic estimate $\hat{F}(t_k)$ by solving

$$\begin{aligned}\mu_C(\hat{A}_{t_k}) &= 1 - \hat{F}(t_k) - \hat{G}(t_{k-1}) + \hat{F}(t_k)\cdot\hat{G}(t_{k-1}) = \frac{n_k - d_k}{n} = k(t_k)\\ &\iff 1 - \hat{G}(t_{k-1}) - \hat{F}(t_k)\big(1 - \hat{G}(t_{k-1})\big) = \frac{n_k - d_k}{n}\\ &\iff 1 - \hat{F}(t_k) = \frac{n_k - d_k}{n}\cdot\frac{1}{(1 - \hat{G}(t_{k-1})}\end{aligned}$$

Using the fact that $1 - \hat{G}(t_{k-1}) = \hat{Z}_{KM}(t_k)$, we find

$$\iff \hat{S}_X(t_k) = \frac{n_k - d_k}{n}\prod_{j=1}^{k-1}\frac{n_j}{n_j - e_j}.$$

By calculating the right-hand side, we conclude $\hat{S}_X(t_k) = \hat{S}_{KM}(t_k)$.

□

[1] See [4], p.131 (4.5) for more details.

5 Implementation

In order to find the Copula-Graphic estimator, we will iteratively solve the following system of linear equations:

$$\mu_C(\hat{A}_{t_i}) - \hat{P}(X > t_i, Y > t_i) = 0 \tag{6}$$
$$\mu_C(\hat{B}_{t_i}) - \hat{P}(X \leq t_i, X < Y) = 0 \tag{7}$$

Ad $\hat{P}(X > t_i, Y > t_i)$ and $\hat{P}(X \leq t_i, X < Y)$:
As discussed, we can directly estimate the two quantities $k(t_i) = \mathbb{P}(X > t_i, Y > t_i)$ and $p_1(t_i) = \mathbb{P}(X \leq t_i, X < Y)$ from the observed sample data. For simplicity, we use the empirical estimates.

$$\hat{P}(X > t_i, Y > t_i) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j > t_i\}} \text{ and } \hat{P}(X \leq t_i, X < Y) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j \leq t_i,\, \delta_j = 1\}}.$$

Remark (estimation of k and p_1)**.**
Note that any consistent estimators of k and p_1 can be used.

Ad $\mu_C(\hat{A}_{t_i})$ and $\mu_C(\hat{B}_{t_i})$:
First, we note that for $\mu_C(\hat{A}_{t_i})$ we have

$$\begin{aligned}\mu_C(\hat{A}_{t_i}) &= \mu_C([\hat{F}(t_i), 1] \times [\hat{G}(t_i), 1]) \\ &= 1 - \hat{F}(t_i) - \hat{G}(t_i) + C(\hat{F}(t_i), \hat{G}(t_i)).\end{aligned}$$

ongo
idea)

In order to compute $\mu_C(\hat{B}_{t_i})$ we need to take a closer look at the structure of $\hat{B}_{t_i}$.

$$\begin{aligned}\hat{B}_{t_i} &= \{(x, y) \mid 0 < x \leq \hat{F}(t_i), \hat{G}\hat{F}^{-1}(x) \leq y \leq 1\} \\ &= \hat{B}_{t_{i-1}} \uplus \left(\hat{B}_{t_i} \setminus \hat{B}_{t_{i-1}}\right) \\ &= \hat{B}_{t_{i-1}} \uplus \underbrace{\{(x, y) \mid \hat{F}(t_{i-1}) < x \leq \hat{F}(t_i), \hat{G}\hat{F}^{-1}(x) \leq y \leq 1\}}_{=: \Delta\hat{B}_{t_i}}.\end{aligned}$$

Assuming that $\mu_C(\hat{B}_{t_{i-1}})$ is known, we only need to estimate the difference $\mu_C(\Delta\hat{B}_{t_i})$.

Ad $\mu_C(\Delta\hat{B}_{t_i})$:
Next, we want to understand the behaviour of $\hat{G}\hat{F}^{-1}$ on the interval $(\hat{F}(t_{i-1}), \hat{F}(t_i)]$.
By construction, $\hat{F}$ is an empirical distribution function and is constant on each interval $[t_{i-1}, t_i)$. So by the definition of $\hat{F}$, we have for any $x \in (\hat{F}(t_{i-1}), \hat{F}(t_i)]$

$$\hat{F}(t) \geq x \iff t \geq t_i.$$

In particular for $(x, y) \in \Delta\hat{B}_{t_i}$

$$x \in (\hat{F}(t_{i-1}), \hat{F}(t_i)] \iff \hat{F}^{-1}(x) = t_i \iff \hat{G}\hat{F}^{-1}(x) = \hat{G}(t_i).$$

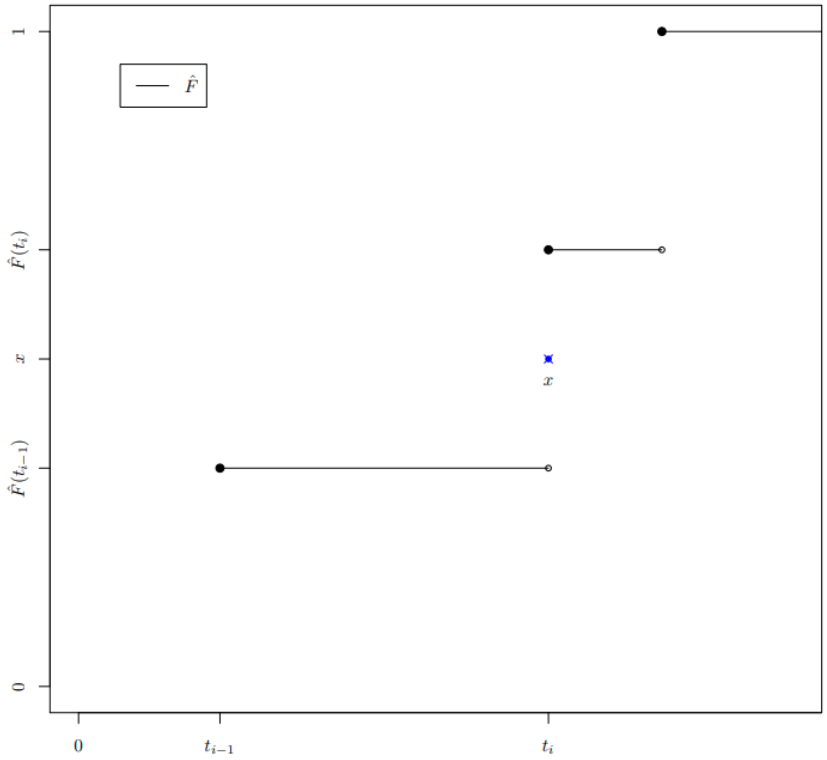


Figure 7: Situation with $\hat{F}(t_{i-1}) < x \leq \hat{F}(t_i)$

It follows that

$$\Delta\hat{B}_{t_i} = \{(x, y) \mid \hat{F}(t_{i-1}) < x \leq \hat{F}(t_i), \hat{G}(t_i) \leq y \leq 1\}.$$

Note that we have

$$\begin{aligned}
\mu_C(\hat{A}_{t_i} \uplus \Delta\hat{B}_{t_i}) &= \mu_C(\{(x, y) \mid \hat{F}(t_{i-1}) < x \leq 1, \hat{G}(t_i) \leq y \leq 1\}) \\
&= \mu_C([\hat{F}(t_{i-1}), 1] \times [\hat{G}(t_i), 1]) \\
&= 1 - \hat{F}(t_{i-1}) - \hat{G}(t_i) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) \\
&= \mu_C(\Delta\hat{B}_{t_i}) + \mu_C(\hat{A}_{t_i}).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mu_C(\Delta\hat{B}_{t_i}) &= 1 - \hat{F}(t_{i-1}) - \hat{G}(t_i) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) \\
&\quad - \big(1 - \hat{F}(t_i) - \hat{G}(t_i) + C(\hat{F}(t_i), \hat{G}(t_i))\big) \\
&= \hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i)).
\end{aligned}$$

We conclude that

$$\begin{aligned}
\mu_C(\hat{B}_{t_i}) &= \mu_C(\hat{B}_{t_{i-1}}) + \hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i)), \\
\mu_C(\hat{A}_{t_i}) &= 1 - \hat{F}(t_i) - \hat{G}(t_i) + C(\hat{F}(t_i), \hat{G}(t_i)).
\end{aligned}$$

C.Longo
(end)

5.1 Iteration to solve for the Copula-Graphic estimates

We iterate through the times $t_i, i = 1, \ldots, n$. In each iteration we solve

$$\overbrace{1 - \hat{F}(t_i) - \hat{G}(t_i) + C(\hat{F}(t_i), \hat{G}(t_i))}^{\mu_C(\hat{A}_{t_i}) =} = \overbrace{\hat{P}(X > t_i, Y > t_i)}^{\hat{k}(t_i) :=}$$

$$\underbrace{\mu_C(\hat{B}_{t_{i-1}}) + \hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i))}_{= \mu_C(\hat{B}_{t_i})} = \underbrace{\hat{P}(X \leq t_i, X < Y)}_{=: \hat{p}_1(t_i)}$$

At time t_i we already know $\hat{F}(t_{i-1}), \hat{G}(t_{i-1})$ and $\hat{B}_{t_{i-1}}$ leaving us with two equations and two unknown variables $\hat{F}(t_i)$ and $\hat{G}(t_i)$. Hence, there exists a unique solution and we can solve for $(\hat{F}(t_i), \hat{G}(t_i))$ using an iterative algorithm.

For $i = 1, 2, \ldots, n$, we find $\hat{F}(t_i)$ and $\hat{G}(t_i)$, by iterating through the following two steps:

Step 1.
Given an initial guess $\hat{F}^{(k)}(t_i)$ for $\hat{F}(t_i)$, we use a bisection algorithm to solve

$$\mu_C(\hat{A}_{t_i}) = 1 - \hat{F}^{(k)}(t_i) - \hat{G}^{(k)}(t_i) + C(\hat{F}^{(k)}(t_i), \hat{G}^{(k)}(t_i)) = \hat{k}(t_i).$$

As result we find an estimate $\hat{G}^{(k)}(t_i)$ for $\hat{G}(t_i)$.

Algorithm 1: Step 1 (Algorithm to solve for $\hat{G}^{(k)}(t_i)$)

Input: $\hat{G}(t_{i-1})$, candidate $\hat{F}^{(k)}(t_i)$ for $\hat{F}(t_i)$
Output: corresponding candidate $\hat{G}^{(k)}(t_i)$ for $\hat{G}(t_i)$
$G_{max} \leftarrow 1$, $G_{min} \leftarrow \hat{G}(t_{i-1})$
while error $> \varepsilon$ **do**
  $G_{trial} \leftarrow \text{mean}(G_{min}, G_{max})$
  $\mu_C^{(k)}(\hat{A}_{t_i}) \leftarrow 1 - \hat{F}^{(k)}(t_i) - G_{trial} + C\big(\hat{F}^{(k)}(t_i), G_{trial}\big)$
  if $\mu_C^{(k)}(\hat{A}_{t_i}) > k(t_i)$ **then**
    $G_{min} \leftarrow G_{trial}$
  else if $\mu_C^{(k)}(\hat{A}_{t_i}) \leq k(t_i)$ **then**
    $G_{max} \leftarrow G_{trial}$
  end
  error $\leftarrow |\mu_C^{(k)}(\hat{A}_{t_i}) - k(t_i)|$
end
$\hat{G}^{(k)}(t_i) \leftarrow G_{trial}$

Step 2.
Next, we need to check if the pair $(\hat{F}^{(k)}(t_i), \hat{G}^{(k)}(t_i))$ satisfies

$$\mu_C(\hat{B}_{t_i}) = \mu_C(\hat{B}_{t_{i-1})} + \hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i)) = \hat{p}_1(t_i).$$

If this equation is satisfied, we accept the pair $(\hat{F}^{(k)}(t_i), \hat{G}^{(k)}(t_i))$, else we adapt the initial guess $\hat{F}^{(k)}(t_i)$ and go back to Step 1.

Algorithm 2: Step 2 (Algorithm for the Copula-Graphic estimator)

Input: $\hat{F}(t_{i-1}), \mu_C(\hat{B}_{t_{i-1}})$

Output: a solution $(\hat{F}(t_i), \hat{G}(t_i))$ for equations (6) and (7)

$F_{max} \leftarrow 1$, $F_{min} \leftarrow \hat{F}(t_{i-1})$

while error $> \varepsilon$ **do**

$\hat{F}^{(k)}(t_i) \leftarrow \text{mean}(F_{min}, F_{max})$

$\hat{G}^{(k)}(t_i) \leftarrow$ Step 1: given $\hat{F}^{(k)}(t_i)$ find corresponding estimate $\hat{G}^{(k)}(t_i)$

$\mu_C^{(k)}(\hat{B}_{t_i}) \leftarrow \mu_C(\hat{B}_{t_{i-1}}) + \hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i))$

if $\mu_C^{(k)}(\hat{B}_{t_i}) > p_1(t_i)$ **then**

  $F_{max} \leftarrow \hat{F}^{(k)}(t_i)$

else if $\mu_C^{(k)}(\hat{B}_{t_i}) \leq p_1(t_i)$ **then**

  $F_{min} \leftarrow \hat{F}^{(k)}(t_i)$

end

error $\leftarrow |\mu_C^{(k)}(\hat{B}_{t_i}) - p_1(t_i)|$

end

$\hat{F}(t_i) \leftarrow \hat{F}^{(k)}(t_i)$, $\hat{G}(t_i) \leftarrow \hat{G}^{(k)}(t_i)$

Theorem 5.1 (Property of the algorithm)**.**
The Copula-Graphic estimate $\hat{F}$ of F jumps at time t_i $(\hat{F}(t_i) > \hat{F}(t_{i-1}))$ if and only if the risk event A was observed at time t_i.

Proof.
Consider the second equation of our system of linear equations at time t_i (i^{th} iteration)

$$\mu_C(\hat{B}_{t_{i-1}} + \hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i)) = \hat{p}(t_i)$$

Assume that at time t_i the risk event A was not observed, i.e. there are no observations of the form $(T_j = t_i, \delta = 1)$, $j = 1, \ldots, n$. In this case we have

$$\hat{p}(t_i) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j \leq t_i,\, \delta_j = 1\}} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{T_j \leq t_{i-1},\, \delta_j = 1\}} = \hat{p}(t_{i-1})$$

Recall that $\mu_C(\hat{B}_{t_{i-1}}) = \hat{p}(t_{i-1})$. So the second equation takes the following form

$$\hat{F}(t_i) - \hat{F}(t_{i-1}) + C(\hat{F}(t_{i-1}), \hat{G}(t_i)) - C(\hat{F}(t_i), \hat{G}(t_i)) = 0.$$

Note that when the algorithm solves this equation for $\hat{F}(t_i)$ in Step 2, the variables $\hat{F}(t_{i-1})$ and $\hat{G}(t_i)$ are fixed input parameters. Hence, we have one equation and one unknown variable, i.e. there exists a unique solution for $\hat{F}(t_i)$. Is is easy to see that this unique solution is given by $\hat{F}_{t_i} = \hat{F}_{t_{i-1}}$. It follows that if the risk event A is not observed at time t_i, then the estimated survival function $\hat{S}$ will not change at time t_i, i.e. $\hat{S}_{t_i} = \hat{S}_{t_{i-1}}$.

□

6 Simulation Studies

6.1 Robustness of the CG estimator w.r.t. assumed Copula

To study the robustness of the Copula-Graphic estimator to misspecification of the copula, the authors generated 10000 samples from the Gamma Frailty copula (see Appendix) with $\tau = 0.5$ (Kendall's τ) and 50% censoring.

For each sample the Copula-Graphic estimator was computed assuming one of the following copulas (see Appendix)

- Gamma copula
- Frank's copula
- Gumbel copula

The parameters are chosen in such way that $\tau = 0.5$. Also computed was the Kaplan-Meier estimator (assumed independence copula).

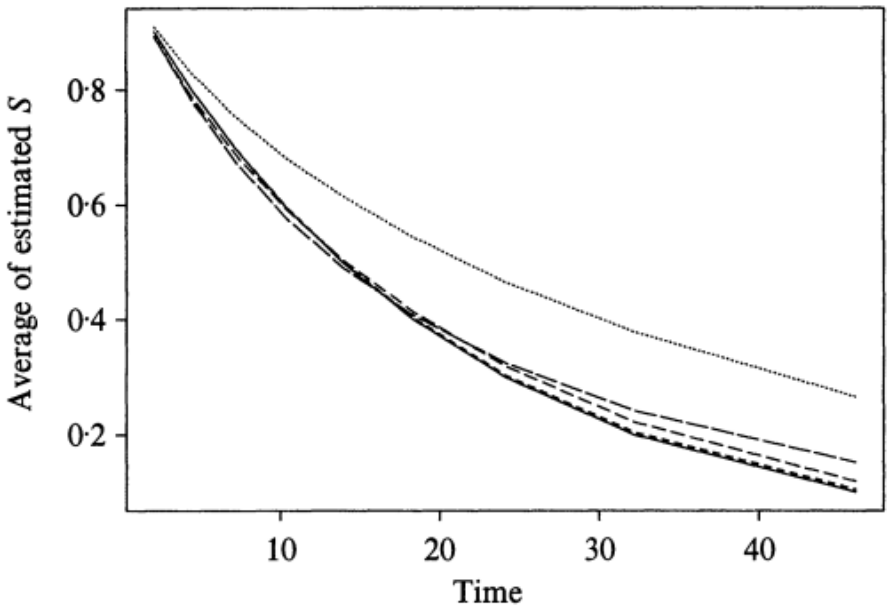


Fig. 3. Robustness of the copula-graphic estimator. Solid line, true survival function; short dashes, copula-graphic estimate using true copula; longer dashes, copula-graphic estimate using Gumbel's copula; longest dashes, copula-graphic estimate using Frank's copula; dotted line, Kaplan–Meier estimate.

Note that there is little difference among the three estimates based on copulas with the correct value of τ, while the Kaplan-Meier estimator tends to overestimate the true survival function by a considerable margin. This figure suggests that the important factor for a sound estimate of the marginal survival function is a reasonable guess at the strength of the association between X and Y and not the functional form of the copula.

7 Case Study: Melanoma data

7.1 Melanoma data (clinical study)

The data that will be used throughout this chapter is available as built-in dataset in the "riskRegression" R package[2].

The study observed a total of 205 patients during the time from 1962 until 1977. At the end 134 were still alive, while 57 had died from cancer and 14 died from other causes. The original dataset contains 12 variables of which we will only consider the following

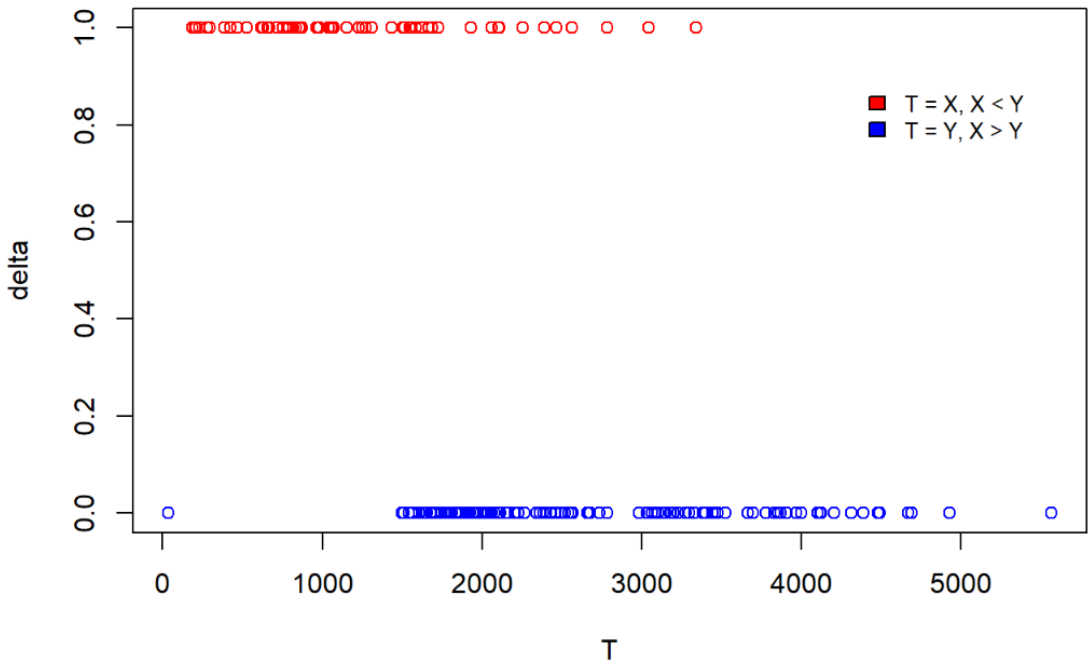
i. time: time-to-event in days from operation (event: "status change"),

ii. status: a numeric with values (0 = censored, 1 = death.melanoma, 2 = death other).

Further, we only consider patients with status 0 or 1, since the Copula-Graphic estimator is designed for the bivariate case of two competing risks. Hence, we observe the two competing events $A = \{\text{died from Melanoma}\}$ and $B = \{\text{censored}\}$ with time-to-event X for A and Y for B.

7.2 Exploratory Data Analysis

The following plot shows the observed sample data (T, δ).

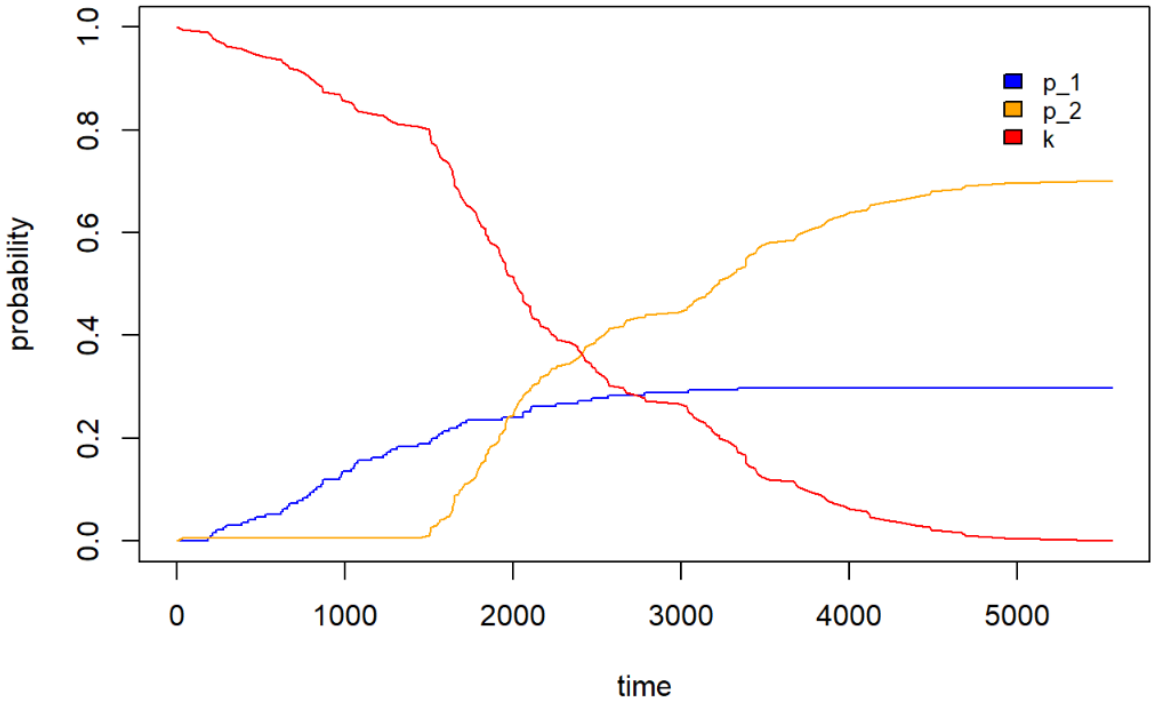


At the beginning of the clinical study there are mostly observations with $X < Y$. These study participants died from Melanoma. At some point in time there appear more and more observations with $X > Y$ and there are only few observations with $X < Y$. An explanation may be that the study participants either defeated cancer and left the study, or they are not satisfied with the results from treatment and therefore leave the study.

[2]See https://rdrr.io/cran/riskRegression/man/Melanoma.html

Given the sample data, we can give empirical estimates for the following quantities

$$k(t) = \mathbb{P}(X > t, Y > t),\ p_1(t) = \mathbb{P}(X \leq t, X < Y) \text{ and } p_2(t) = \mathbb{P}(Y \leq t, Y < X).$$

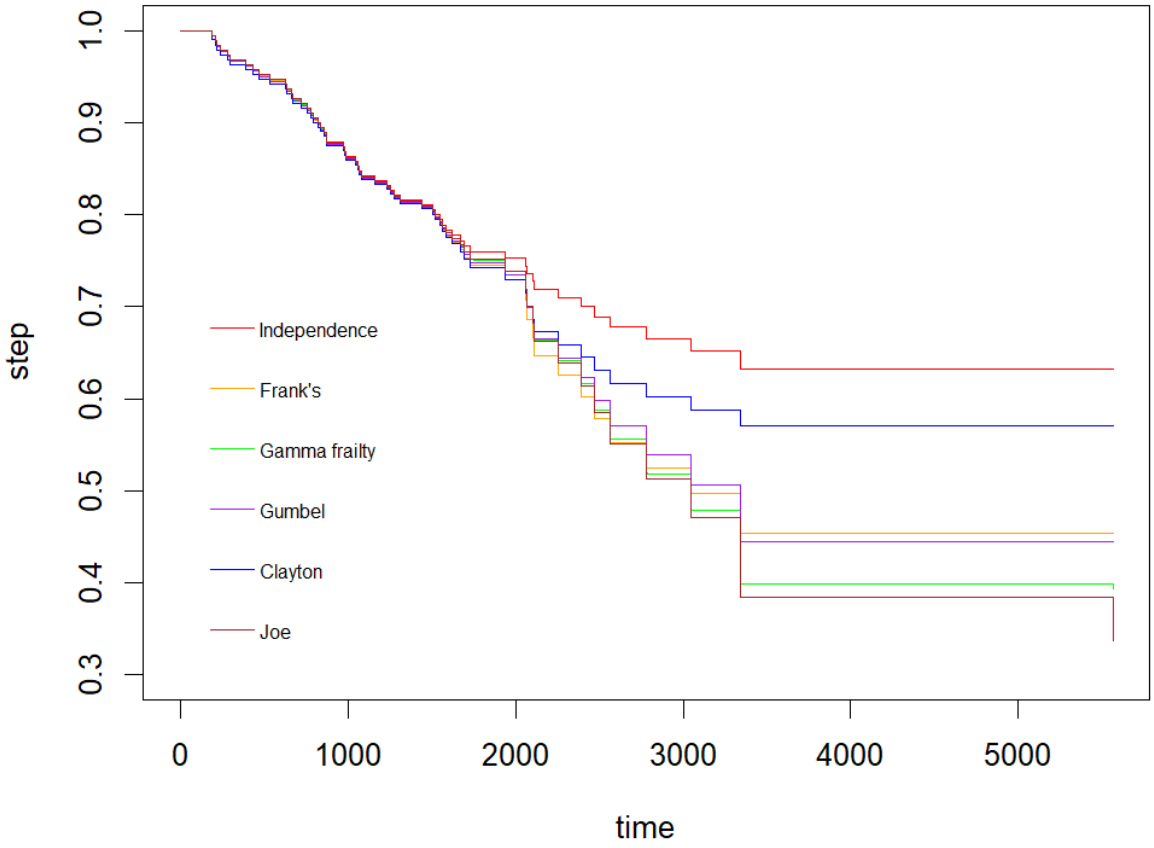


- $k(t)$: $\{X > t, Y > t\} \longrightarrow$ joint survival, i.e. at time t the study participant is still part of the study
- $p_1(t)$: $\{X \leq t, X < Y\} \longrightarrow$ the patient died from Melanoma at a time before t while still being a participant of the study
- $p_2(t)$: $\{Y \leq t, Y < X\} \longrightarrow$ the patient left the study and may or may not die from Melanoma at an unobserved time after t

7.3 Simulation studies on clinical study data

For the simulation study, we calculate the Copula-Graphic estimate $\hat{F}$ of F using different assumed copulas. The applied copulas are listed here (see Appendix)

- independence copula (Kaplan-Meier estimate)
- Frank copula
- Gamma frailty copula
- Gumbel copula
- Clayton copula
- Joe copula



We can see that the estimates of F are very similar for times t in $[0, 1600]$. This is because until time $t = 1500$ there are close to no observations where participants leave the study (censoring event). Hence, the assumed dependency structure between X and Y does not come into play yet. Starting from time $t = 1500$ there are only few observed deaths and at the same time many participants are leaving the study. So the assumed dependency structure plays an increasingly important role.

If we assume independence between X and Y (independence copula), we observe that the survival function does not differ much from the estimates with other assumed dependency structures in the time interval $[0, 1600]$. But after $t = 1500$, when a lot of the participants leave the study (censoring the time X), there is a big difference in the behaviour of the estimator based on the independence copula compared to the other estimators. To be more concrete, the estimate based on independence overestimates the survival probability to times after $t = 1500$. This is because the other copulas imply a dependency of X and Y and hence if many participants leave the study, these estimates take into account that some of those who left the study would have died later on, which then decreases the probability of survival.

8 Appendix

8.1 Copula families

Definition 8.1 (Gamma Frailty Copula)**.**

$$C(x,y) = x + y - 1 + \Big(\Big(\frac{1}{1-x}\Big)^{(\alpha-1)} + \Big(\frac{1}{1-y}\Big)^{(\alpha-1)} - 1\Big)^{-\frac{1}{\alpha-1}}, \quad \alpha \geq 1.$$

Definition 8.2 (Frank's Copula)**.**

$$C(x,y) = -\frac{1}{\theta}\log\Big(1 + \frac{(\exp(-\theta x)-1)(\exp(-\theta y)-1)}{\exp(-\theta)-1}\Big), \quad \theta \neq 0.$$

Definition 8.3 (Gumbel Copula)**.**

$$C(x,y) = \exp\Big(-\big((-\log(x))^{\theta} + (-\log(x))^{\theta}\big)^{\frac{1}{\theta}}\Big), \quad \theta \geq 1, \quad \theta \in [1,\infty)$$

Definition 8.4 (Joe Copula)**.**

$$C(x,y) = 1 - \big((1-x)^{\theta} + (1-y)^{\theta} - (1-x)^{\theta}(1-y)^{\theta}\big)^{\frac{1}{\theta}}, \quad \theta \geq 1$$

References

[1] Simeon M. Berman. "Note on Extreme Values, Competing Risks and Semi-Markov Processes". In: *The Annals of Mathematical Statistics* 34 (1963).

[2] James B. Robertson and V. R. R. Uppuluiu. "A generalized Kaplan-Meier estimator". In: *The Annals of Statistics* 12 (1984).

[3] Zheng. "On the use of copulas in dependent competing risk theory". In: (1992).

[4] Ming Zheng and John Klein. "Estimates of Marginal Survival for Dependent Competing Risks Based on an Assumed Copula". In: *Biometrika* 82 (1995).