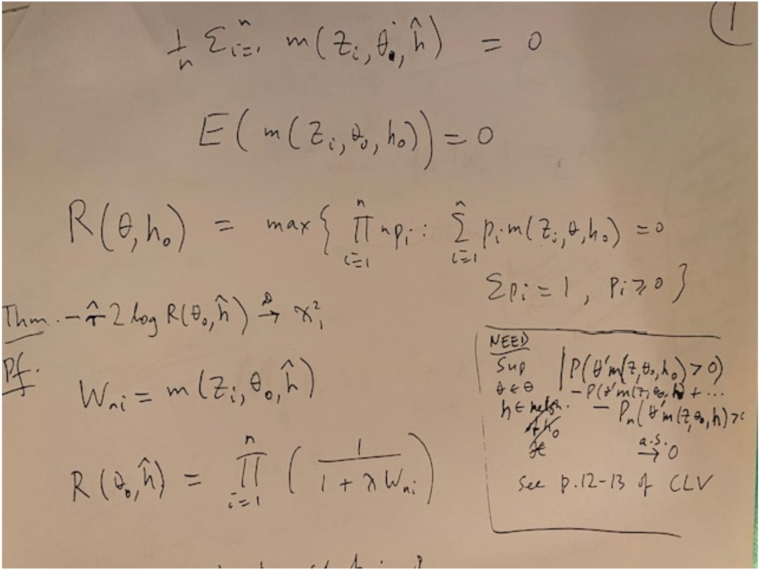
Copula based Cox proportional hazards models for dependent censoring

Ingrid Van Keilegom

December 4, 2023

KU LEUVEN

Idea paper 1:



1

$$\frac{1}{n}\sum_{i=1}^{n} m(Z_i,\theta_0,\hat{h}) = 0$$

$$E\left(m(Z_i,\theta_0,h_0)\right) = 0$$

$$R(\theta,h_0) = \max\left\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i\, m(Z_i,\theta,h_0) = 0, \ \sum p_i = 1,\ p_i \geq 0\right\}$$

Thm. $-\hat{\tau}\, 2\log R(\theta_0,\hat{h}) \xrightarrow{d} \chi^2_1$

Pf. $W_{ni} = m(Z_i,\theta_0,\hat{h})$

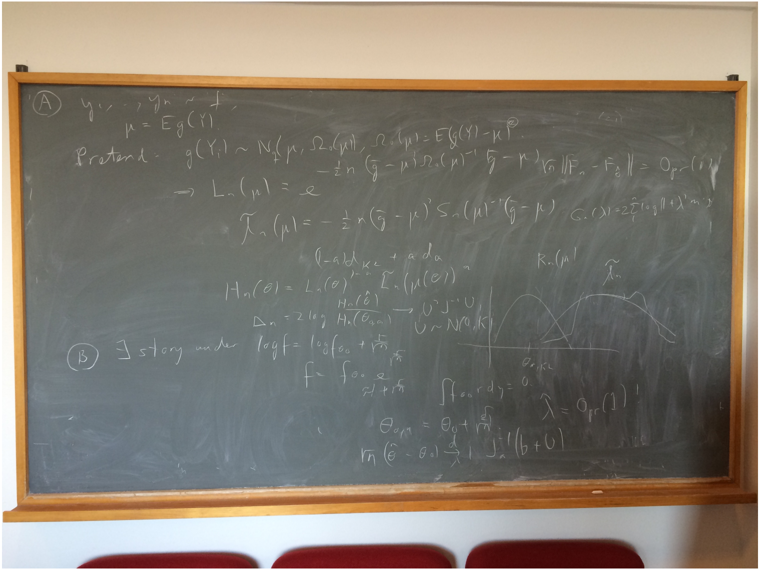
$$R(\theta_0,\hat{h}) = \prod_{i=1}^{n}\left(\frac{1}{1+\lambda W_{ni}}\right)$$

NEED

$$\sup_{\substack{\theta\in\Theta \\ h\in \text{nbhd of } h_0}} \left| P(\theta' m(Z,\theta_0,h_0) > 0) - P(\theta' m(Z,\theta_0,h)) + \dots - P_n(\theta' m(Z,\theta_0,h) > 0)\right| \xrightarrow{a.s.} 0$$

See p.12-13 of CLV

Idea paper 2:



Oslo, September 2014

Oslo, May 2016

Brussels, June 2017

Consider a survival time T subject to random right censoring

$\Rightarrow$ We observe

$$Y = \min(T, C) \text{ and } \Delta = I(T \leq C),$$

where C is a censoring variable

In many situations we observe either T or C, but not both

$\Rightarrow$ Relation between T and C not identifiable nonparametrically (Tsiatis, 1975)

$\Rightarrow$ $F_{T,C}$ not identifiable based on law of (Y, Δ)

$\Rightarrow$ Also F_T not identifiable

To overcome this, it is commonly assumed that T and C are stochastically independent, which solves the identification problem

But is this independence assumption always satisfied in practice ?

Independence between T and C cannot be tested, but the context of a study can give useful insight into the validity of this assumption

Independence of T and C is satisfied if

- ◇ Administrative censoring: individuals alive at the end of the study are censored
  $\Rightarrow$ Censoring is unrelated to survival time
  $\Rightarrow$ Independence assumption makes sense
- ◇ Censoring happens for other reasons that are completely unrelated to the event of interest
- ◇ Many other contexts

Independence of T and C might be doubtful in

- Medical studies : Patients may withdraw from the study
  - because their condition is deteriorating or because they are showing side effects which need alternative treatments (positive relation between T and C)
  - because their health condition has improved and so they no longer follow the treatment (negative relation between T and C)
- Unemployment studies : Unemployed people with low chances on the job market could decide to go abroad to improve their chances, leading to censoring times that depend on the duration of unemployment
- Transplant studies : Often the length of time a patient has to wait before he gets transplanted (C) depends on his/her medical condition, so on his time to death (T)

What happens if independence is assumed when T and C are in reality correlated ?

Consider

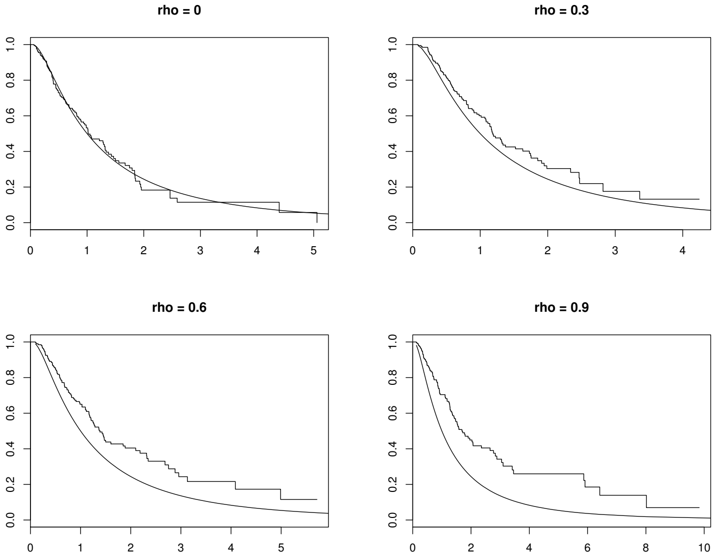
$$(\log T, \log C) \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

where $\rho = 0, \pm 0.3, \pm 0.6$ or ± 0.9

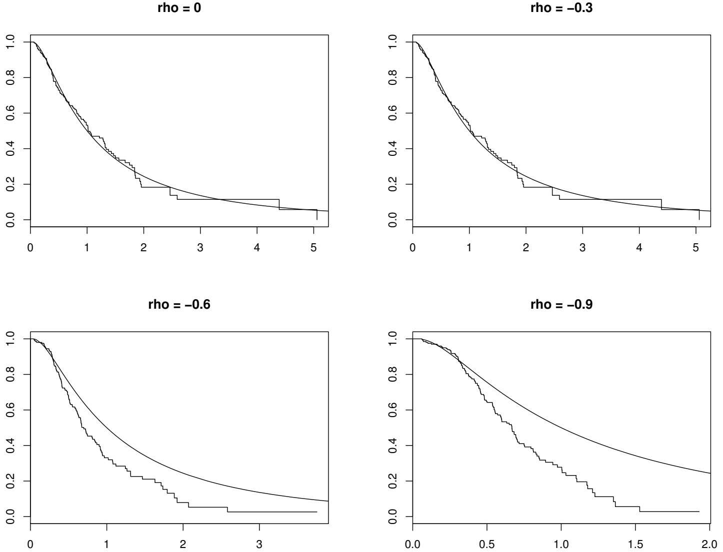
Further, let $Y = \min(T, C)$ and $\Delta = I(T \leq C)$

For an arbitrary sample of size $n = 200$, we calculate

- ◇ the true survival function $S(t)$ of $T \sim \exp(N(0, 1))$
- ◇ the Kaplan-Meier estimator $\hat{S}(t)$ (which assumes $T \perp\!\!\!\perp C$)



$\Rightarrow$ The larger ρ, the more the Kaplan-Meier estimator lies above the true survival function



⇒ The smaller ρ, the more the Kaplan-Meier estimator lies below the true survival function

This talk is NOT about

- Competing risks:
  Choice between dependent censoring and competing risks often depends on the research question

  More

- Informative censoring:
  E.g. Koziol-Green model, models for T and C with common parameters, ...

  More

- Dependent censoring caused by observed covariates:
  We suppose that even after conditioning on observed covariates, T and C are still dependent

What's in a name ? The above concepts are often confused conceptually with the concept of dependent censoring

Literature on dependent censoring

The most popular approach is based on copulas:

What is a copula ?

A bivariate distribution function on $[0,1] \times [0,1]$ with uniform margins

Sklar's theorem

Suppose $X \sim F$, $Y \sim G$
If F and G are continuous, there exists a unique copula $\mathcal{C}$ such that

$$P(X \leq x, Y \leq y) = \mathcal{C}\big(F(x), G(y)\big)$$

If X and Y are independent, then

$$P(X \leq x, Y \leq y) = F(x)G(y) = \mathcal{C}\big(F(x), G(y)\big)$$

with $\mathcal{C}(u, v) = uv$, called the independence copula

Approaches based on copulas:

- Zheng and Klein (1995):
  - ◇ Modelling of the bivariate distribution of T and C by means of a fully known copula function :

$$P(T > t, C > c) = \mathcal{C}(S_T(t), S_C(c))$$

  - ◇ Nonparametric estimation of the survival function S_T under this copula model $\Rightarrow$ extension of Kaplan-Meier
- Rivest and Wells (2001): special case of Archimedean copulas
- Braekers and Veraverbeke (2005), Huang and Zhang (2008), Sujica and VK (2015, 2018), Emura and Chen (2018): extensions to regression models

But: All approaches assume that the copula is fully known with non- or semiparametric margins

Relaxation of the known copula assumption:

- Czado and VK, 2023 (Biometrika)
  - First paper to show identifiability of copula model for dependent censoring without assuming that copula is fully known
  - Approach lies the foundations of this approach, but is limited to parametric model without covariates

- Deresa, Antonio and VK, 2022 (Insur. Math. Econ.)
  - Parametric margins that depend on covariates
  - Parametric copula independent of covariates
  - Dependent censoring and truncation
  - Broader marginal parametric models
  - Actuarial application

Question we want to address is

How to allow for semiparametric models and at the same time avoid the known-copula-assumption ?

We will do that under the following model framework:

- T follows semiparametric Cox model
- C follows parametric regression model
- Copula is parametric and independent of covariates

Reference:

Deresa, N.W. and VK (2023). Copula based Cox proportional hazards models for dependent censoring. *J. Amer. Statist. Assoc.* (to appear).

Consider

- ◇ a survival time T
- ◇ a vector of covariates X

Suppose that $T|X$ follows a Cox proportional hazards model:

$$F_{T|X}(t|x) = P(T \leq t|X = x) = 1 - \exp\{-\Lambda(t)e^{x^\top \beta}\},$$

for some - unspecified baseline cumulative hazard Λ
- vector of regression parameters β

Suppose that instead of observing T we observe

$$Z = \min(T, C, A), \quad \Delta_1 = I(Z = T), \quad \Delta_2 = I(Z = C),$$

where - C is a dependent censoring time
- A is an independent censoring time

We suppose the following models for C and A:

- For some parameter space H, and some vector of covariates W,

$$F_{C|W} \in \{F_{C|W,\eta} : \eta \in H\}$$

- The law of A is unspecified
- $A \perp\!\!\!\perp (T, C)|(X, W)$ and $A \perp\!\!\!\perp (X, W)$

Finally, to model the dependence between T and C we use a copula model:

$$P(T \leq t, C \leq c|X = x, W = w) = \mathcal{C}(F_{T|X}(t|x), F_{C|W}(c|w)),$$

where $\mathcal{C} \in \{\mathcal{C}_\gamma : \gamma \in \Gamma\}$ for some parameter space Γ

This model can be extended to more complex models:

- ⋄ Semiparametric model for C (e.g. Cox model)
- ⋄ Copula parameters depending on covariates
- ⋄ Extensions of Cox model (e.g. transformation model)
- ⋄ More complex censoring schemes

⇒ We will discuss some of these extensions at the end

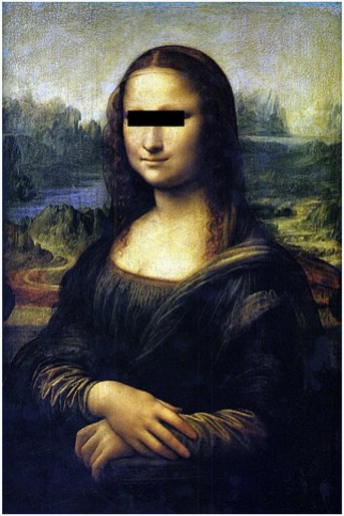
It what follows, we will need

$$h_{T|C}(u|v) = \frac{\partial}{\partial v}\mathcal{C}(u, v), \quad \text{and} \quad h_{C|T}(v|u) = \frac{\partial}{\partial u}\mathcal{C}(u, v)$$

Then,

$$P(T \le t | C = c, X = x, W = w) = h_{T|C}(F_{T|X}(t|x) | F_{C|W}(c|w))$$

Is this model identifiable ? Under which conditions ?



With identifiability we mean that any two different sets of parameters give different joint distributions of $(Z, \Delta_1, \Delta_2, X, W)$

We will show identifiability under the following conditions (C1)-(C5):

(C1) The matrices $\text{Var}(X)$ and $\text{Var}(W)$ have full rank

(C2) The vectors X and W contain at least one continuous variable

(C3) For all $\eta_1, \eta_2 \in H$, we have:

$$\lim_{t \to 0} \frac{f_{C|W,\eta_1}(t|w)}{f_{C|W,\eta_2}(t|w)} = 1 \quad \text{for all } w \iff \eta_1 = \eta_2$$

Lemma

Condition (C3) is satisfied for the families of log-normal, log-Student-t, Weibull, and log-logistic densities.

(C4) For all γ, all $\zeta = (\beta, \Lambda)$ and all η,

$$\lim_{t \to 0} h_{T|C,\gamma}(F_{T|X,\zeta}(t|x)|F_{C|W,\eta}(t|w)) = 0 \text{ for all } (x, w)$$

The same holds true for $h_{C|T,\gamma}$

Lemma

Condition (C4) is satisfied by

(1) *the Frank copula, independently of the marginal distributions*

(2) *the Gumbel copula if for all x, w, ζ, η,*

$$0 < \lim_{t \to 0} \frac{\log F_{T|X,\zeta}(t|x)}{\log F_{C|W,\eta}(t|w)} < \infty$$

(3) *the Gaussian copula if for all $x, w, \zeta, \eta, \gamma$,*

$$\lim_{t \to 0} [\Phi^{-1}(F_{T|X,\zeta}(t|x)) - \gamma \Phi^{-1}(F_{C|W,\eta}(t|w))] = -\infty$$

$$\lim_{t \to 0} [\Phi^{-1}(F_{C|W,\eta}(t|w)) - \gamma \Phi^{-1}(F_{T|X,\zeta}(t|x))] = -\infty$$

Remark:

- ◇ Gumbel copula: Note e.g. that

$$\lim_{t\to 0} \frac{\log F_{T|X,\zeta}(t|x)}{\log F_{C|W,\eta}(t|w)} = \frac{\rho_T(x)}{\rho_C(w)} \in (0,\infty)$$

if $\quad T|X = x \sim \text{Weibull}(\lambda_T(x), \rho_T(x))$
$\quad\quad C|W = w \sim \text{Weibull}(\lambda_C(w), \rho_C(w))$

- ◇ Gaussian copula: (C4) is satisfied for many common margins (numerical verification)

(C5) For all $\gamma_k, \zeta_k = (\beta, \Lambda_k), \eta$ ($k = 1, 2$) that are such that $\lim_{t\to 0} \lambda_1(t)/\lambda_2(t) = 1$, we have

$$\lim_{t\to 0} \frac{c_{\gamma_1}(F_{T|X,\zeta_1}(t|x), F_{C|W,\eta}(t|w))}{c_{\gamma_2}(F_{T|X,\zeta_2}(t|x), F_{C|W,\eta}(t|w))} = 1 \text{ for all } (x, w) \iff \gamma_1 = \gamma_2,$$

where c_γ denotes the copula density

Lemma

Condition (C5) is satisfied for the Frank, Gumbel and Gaussian copulas

Theorem

Assume that conditions (C1)-(C5) hold true. Then, our model is identifiable.

Some remarks :

- Case of the Clayton copula
- Survival copulas are also possible
- Conditions are sufficient but not necessary

Assume that we have an i.i.d. sample $\{(Z_i, \Delta_{1i}, \Delta_{2i}, X_i, W_i)\}_{i=1}^n$ of $(Z, \Delta_1, \Delta_2, X, W)$, where

$$Z = \min(T, C, A), \quad \Delta_1 = I(Z = T), \quad \Delta_2 = I(Z = C)$$

$\Rightarrow$ The likelihood for $\theta = (\gamma, \beta, \eta)$ is given by

$$
\begin{aligned}
& L(\theta, \Lambda) \\
&= \prod_{i=1}^n \Big[\lambda(Z_i) e^{X_i^\top \beta} \exp\{-\Lambda(Z_i) e^{X_i^\top \beta}\} \\
&\qquad\qquad \times \big\{1 - h_{C|T,\gamma}(F_{C|W,\eta}(Z_i|W_i)|F_{T|X,\zeta}(Z_i|X_i))\big\}\Big]^{\Delta_{1i}} \\
&\quad \times \Big[f_{C|W,\eta}(Z_i|W_i)\{1 - h_{T|C,\gamma}(F_{T|X,\zeta}(Z_i|X_i)|F_{C|W,\eta}(Z_i|W_i))\}\Big]^{\Delta_{2i}} \\
&\quad \times \Big[\tilde{\mathfrak{C}}_\gamma\Big\{F_{T|X,\zeta}(Z_i|X_i), F_{C|W,\eta}(Z_i|W_i)\Big\}\Big]^{(1-\Delta_{1i})(1-\Delta_{2i})}
\end{aligned}
$$

since the density and distribution of A can be omitted from the likelihood

Main idea:

Direct maximization of this likelihood is challenging since it involves the unknown function Λ

$\Rightarrow$ We estimate θ by replacing Λ in the likelihood with a nonparametric estimator $\hat{\Lambda}(\cdot, \theta)$ for fixed θ

$\Rightarrow$ θ is then estimated by solving the score equation derived from the pseudo-likelihood $L(\theta, \hat{\Lambda}(\cdot, \theta))$

We will use martingale ideas to construct $\hat{\Lambda}(\cdot, \theta)$. For all i, let

- $N_i(z) = I(Z_i \leq z, \Delta_{1i} = 1)$ and $Y_i(z) = I(Z_i \geq z)$
- $\tau_0 =$ finite maximum follow-up time

and define the conditional crude hazard rate $\lambda^{\#}(z|X, W)$:

$$\lambda^{\#}(z|X, W) = \frac{-\frac{\partial}{\partial u} P(T \geq u, C \geq z|X, W)|_{u=z}}{P(T \geq z, C \geq z|X, W)}$$

Then,

$$M_i(z) = N_i(z) - \int_0^z Y_i(s)\lambda^{\#}(s|X_i, W_i)ds$$

is a martingale with respect to the filtration

$$\mathcal{F}_z^i = \sigma\{Y_i(s), N_i(s), X_i, W_i;\ 0 \le s \le z \le \tau_0\}$$

Under the general parametric copula model, we have that

$$M_i(z) = N_i(z) - \int_0^z Y_i(s) \exp(\psi_i(s, \theta_0, \Lambda_0)) d\Lambda_0(s),$$

for a certain function ψ_i, where $\theta_0 = (\gamma_0, \beta_0, \eta_0)$

$\Rightarrow$ We estimate Λ for a given θ by solving the estimating equation

$$\sum_{i=1}^{n} \{dN_i(z) - Y_i(z) \exp(\psi_i(z, \theta, \Lambda)) d\Lambda(z)\} = 0 \quad (0 \le z \le \tau_0)$$

⇒ The estimator $\hat{\Lambda}(\cdot, \theta)$ is a nondecreasing step function with jumps only at the observed survival times, denoted by $z_1 < \cdots < z_K < \infty$

But: it involves a complex iterative optimization process

⇒ We propose an alternative estimator that is simpler to compute, and that consists in replacing $\psi_i(z, \theta, \Lambda)$ by $\psi_i(z-, \theta, \Lambda)$ in the estimating equation:

$$\begin{aligned}
\Delta\hat{\Lambda}(z_k, \theta) &= \hat{\Lambda}(z_k, \theta) - \hat{\Lambda}(z_{k-1}, \theta) \\
&= \frac{\sum_{i=1}^n dN_i(z_k)}{\sum_{i=1}^n Y_i(z_k) \exp\{\psi_i(z_{k-1}, \theta, \hat{\Lambda})\}}
\end{aligned}$$

Note that $\Delta\hat{\Lambda}(z_k, \theta)$ depends on $\hat{\Lambda}(z_j, \theta)$ for $j = 1, \ldots, k-1$

⇒ Avoids iterative optimization scheme for estimating Λ

We now estimate θ by replacing Λ by $\hat{\Lambda}(\cdot, \theta)$ in the likelihood

$$L(\theta, \Lambda) = \prod_{i=1}^{n} g_{\theta,\Lambda}(Z_i, \Delta_{1i}, \Delta_{2i} | X_i, W_i)$$

and setting the derivative with respect to θ to zero

$\Rightarrow$ This gives the following estimating equation:

$$U_n(\theta, \hat{\Lambda}(\cdot, \theta)) = n^{-1} \sum_{i=1}^{n} U(Z_i, \Delta_{1i}, \Delta_{2i}, \theta, \hat{\Lambda}(\cdot, \theta)) = 0,$$

where

$$U(Z_i, \Delta_{1i}, \Delta_{2i}, \theta, \Lambda) = \frac{\partial}{\partial \theta} \log g_{\theta,\Lambda}(Z_i, \Delta_{1i}, \Delta_{2i} | X_i, W_i)$$

Finally, $\hat{\theta}$ is defined as a solution of this score equation

What happens in the special case of the independence copula $\mathcal{C}_\gamma(u, v) = uv$?

$\Rightarrow \hat{\Lambda}$ cancels out from the formula of $\psi_i(z_{k-1}, \theta, \hat{\Lambda})$

$\Rightarrow \hat{\Lambda}(\cdot, \theta)$ reduces to the Breslow estimator of the cumulative hazard function in the Cox model (Breslow, 1974)

$\Rightarrow \hat{\theta}$ reduces to the partial likelihood estimator of Cox (Cox, 1972)

$\Rightarrow$ Proposed estimator of θ is extension of the partial likelihood estimator to the case of dependent censoring

Lemma

(i) *Consistency and rate of convergence of* $\hat{\Lambda}(\cdot, \theta)$*:*

$$\sup_{\theta \in \Theta, 0 \leq z \leq \tau_0} |\hat{\Lambda}(z, \theta) - \Lambda_0(z, \theta)| = O_p(n^{-1/2})$$

(ii) *Iid representation of* $\hat{\Lambda}(\cdot, \theta_0) - \Lambda_0(\cdot)$*:*

$$\hat{\Lambda}(z, \theta_0) - \Lambda_0(z) = \frac{1}{A(z)} \frac{1}{n} \sum_{i=1}^{n} \int_0^z \frac{A(s)}{B(s)} \, dM_i(s) + R_n(z),$$

where $\sup_{0 \leq z \leq \tau_0} |R_n(z)| = o_p(n^{-1/2})$

(iii) *Consistency of* $(\partial / \partial \theta) \hat{\Lambda}(\cdot, \theta_0)$*:*

$$\left. \frac{\partial \hat{\Lambda}(z, \theta)}{\partial \theta} \right|_{\theta = \theta_0} = \frac{1}{A_1(z)} \int_0^z \frac{A_1(s)}{B(s)} dD(s) + o_p(1),$$

for every $z \in [0, \tau_0]$

Theorem

(i) *Consistency of $\hat{\theta}$:*

$$\hat{\theta} \xrightarrow{P} \theta_0$$

(ii) *Asymptotic normality of $\hat{\theta}$:*

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}\{0, \Sigma_1^{-1}\Sigma_2(\Sigma_1^{-1})^\top\}$$

Remarks:

- Proof is based on Chen, Linton, VK (2003) containing primitive conditions for consistency and asymptotic normality of semiparametric Z-estimators
- Asymptotic variance has explicit but complex formula
  $\Rightarrow$ Bootstrap will be used instead

Scenario 1: Comparison with independence model

- Frank copula with Kendall's $\tau = 0.2$, 0.4 or 0.8
- Cox model for T:

$$F_{T|X}(t|x) = 1 - \exp\Big(- \Lambda(t) e^{\beta_1 x_1 + \beta_2 x_2} \Big)$$

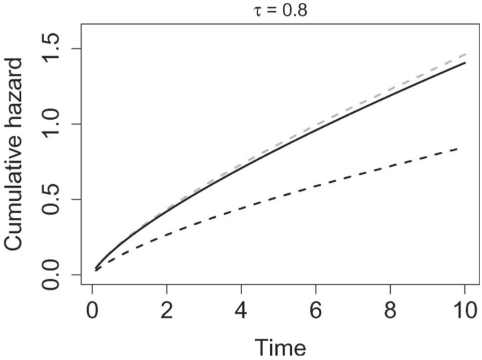
  with $\Lambda(t) = 0.25 t^{3/4}$, $\beta_1 = 0.45$ and $\beta_2 = 1$
- Weibull model for C:

$$F_{C|X}(t|x) = 1 - \exp\Big(- \exp\Big(\frac{\log(t) - (\eta_0 + \eta_1 x_1 + \eta_2 x_2)}{\sigma} \Big) \Big),$$

  with $\eta_0 = 1.35, \eta_1 = 0.3, \eta_2 = 1$ and $\sigma = 1$
- $X_1 \sim \text{Bern}(0.5)$, $X_2 \sim \mathcal{N}(0,1)$, and $X_1 \perp\!\!\!\perp X_2$
- $A \sim U[0, 15]$ and $A \perp\!\!\!\perp (T, C, X_1, X_2)$
- 1000 data sets of size $n = 500$ are used

$\Rightarrow$ We have approximately 45% T, 40% C and 15% A

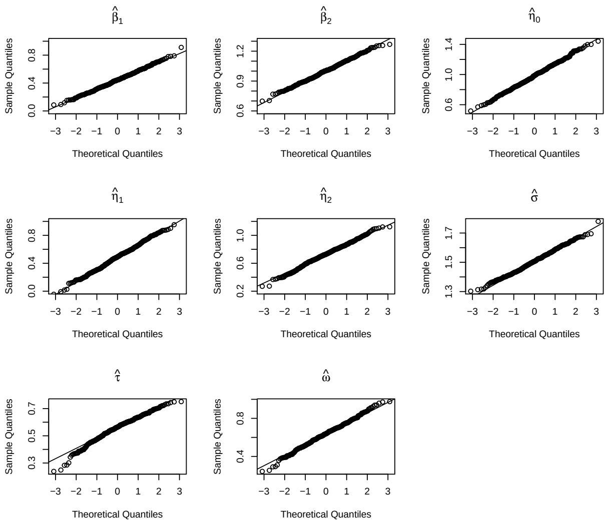
Average of the estimated cumulative hazard functions:



Frank copula: dashed grey line
Independence copula: dashed black line
True cumulative hazard function: solid line

	$\tau = 0.2$			τ = 0.4			τ = 0.8		
	Bias	ESD	RMSE	Bias	ESD	RMSE	Bias	ESD	RMSE
	Frank copula								
β_1	-0.010	0.134	0.134	-0.014	0.131	0.131	-0.024	0.126	0.129
β_2	-0.003	0.098	0.098	-0.004	0.099	0.099	-0.013	0.102	0.103
η_0	-0.005	0.139	0.139	-0.006	0.123	0.123	-0.022	0.113	0.115
η_1	0.001	0.136	0.136	0.004	0.125	0.125	0.013	0.110	0.111
η_2	-0.002	0.118	0.118	-0.002	0.109	0.109	-0.012	0.099	0.100
σ	-0.002	0.052	0.052	-0.001	0.052	0.052	0.003	0.051	0.051
τ	0.012	0.112	0.112	0.010	0.090	0.090	0.008	0.037	0.038
	Independence copula								
β_1	0.024	0.135	0.137	0.058	0.137	0.149	0.124	0.141	0.188
β_2	0.081	0.085	0.118	0.167	0.087	0.188	0.327	0.092	0.340
η_0	0.165	0.111	0.199	0.314	0.109	0.333	0.520	0.110	0.532
η_1	0.052	0.140	0.150	0.100	0.137	0.170	0.169	0.133	0.215
η_2	0.129	0.095	0.160	0.248	0.095	0.265	0.415	0.101	0.427
σ	0.001	0.055	0.055	-0.013	0.055	0.057	-0.082	0.053	0.098

Normality of the estimators:



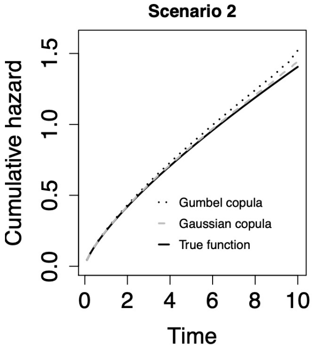
$\hat{\omega} =$ Fisher's Z transformation of $\hat{\tau}$

Estimation of the variance and 95% coverage rates:

Par.	Bias	ESD	BSE	RMSE	CR
β_1	-0.006	0.130	0.136	0.130	0.959
β_2	0.002	0.100	0.104	0.100	0.958
η_0	-0.021	0.160	0.172	0.161	0.968
η_1	-0.011	0.178	0.173	0.179	0.950
η_2	-0.024	0.148	0.156	0.150	0.948
σ	0.008	0.076	0.078	0.077	0.958
τ	0.019	0.084	0.085	0.086	0.928

Scenario 2: Sensitivity to misspecification of the copula structure

- Same model as for Scenario 1 except that Gumbel and Gaussian copulas are used to estimate the model
- Average of the estimated cumulative hazard functions:



$\Rightarrow$ Findings similar to those in Huang and Zhang (2008)

	$\tau = 0.2$			τ = 0.4			τ = 0.8		
	Bias	ESD	RMSE	Bias	ESD	RMSE	Bias	ESD	RMSE
	Gumbel copula								
β_1	-0.009	0.139	0.140	-0.021	0.138	0.139	-0.019	0.126	0.128
β_2	0.006	0.104	0.104	-0.003	0.108	0.108	-0.012	0.099	0.100
η_0	-0.005	0.158	0.158	-0.022	0.141	0.142	-0.026	0.112	0.115
η_1	0.002	0.139	0.139	0.000	0.129	0.129	0.006	0.112	0.112
η_2	-0.004	0.135	0.135	-0.018	0.124	0.125	-0.019	0.100	0.102
σ	-0.014	0.052	0.054	-0.014	0.052	0.053	0.006	0.051	0.052
τ	-0.001	0.130	0.130	0.013	0.109	0.110	0.000	0.036	0.036
	Gaussian copula								
β_1	-0.013	0.135	0.135	-0.018	0.132	0.133	-0.008	0.124	0.125
β_2	-0.011	0.107	0.107	-0.016	0.105	0.106	-0.000	0.098	0.098
η_0	-0.018	0.159	0.160	-0.021	0.132	0.133	-0.001	0.107	0.107
η_1	-0.002	0.139	0.139	-0.000	0.127	0.127	0.012	0.112	0.113
η_2	-0.010	0.133	0.133	-0.012	0.117	0.117	0.007	0.097	0.097
σ	0.004	0.053	0.054	0.012	0.053	0.054	0.020	0.053	0.056
τ	0.022	0.140	0.142	0.014	0.096	0.097	-0.047	0.043	0.064

Scenario 3: Goodness-of-fit tests for Cox/copula model

- ◇ The idea is to construct a test statistic from the L_2 distance between a model based and a nonparametric estimator of the distribution of $R = \min(T, C)$
- ◇ Model based estimator: Can be derived from the expressions of $\hat{F}_T$ and $\hat{F}_C$
- ◇ Nonparametric estimator: Since $R \perp\!\!\!\perp A$, a regular Kaplan-Meier estimator of F_R can be used
- ◇ Bootstrap is used under H_0 to approximate the rejection rates

Three cases:

- Case 1: Correctly specified model
- Case 2: Model for C misspecified
- Case 3: Regression functions for T and C misspecified

Rejection rates:

n	Case	5%	10%
500	1	0.038	0.078
	2	0.504	0.674
	3	0.334	0.430
1000	1	0.058	0.122
	2	0.938	0.976
	3	0.651	0.765

What we are currently working on:

◇ Extension to the case where both T and C follow a semiparametric transformation model

◇ Dependent censoring in cure models

◇ Dependent censoring and confounding based on semiparametric Cox model for T

◇ Quantile regression under dependent censoring

◇ Investigation of partial identification results

◇ Random effects approach to handle dependent censoring

Main reference:

Deresa, N.W. and VK (2023). Copula based Cox proportional hazards models for dependent censoring. *Journal of the American Statistical Association (to appear)*, DOI: 10.1080/01621459.2022.2161387

Example: Staphylococcus infection (Geskus, 2016)

- ◇ Of interest : Time to infection during in-hospital stay
- ◇ How to deal with patients that are discharged without infection ?

(1) Biological question : What would happen if everyone stayed in hospital ? (relevant to compare infection risk with other hospitals)

$\Rightarrow$ Use marginal distribution (with discharge considered as censoring event)

$\Rightarrow$ Leads to dependent censoring

(2) Clinical question : What percentage of patients gets infected while staying in hospital, and when do they get infected ?

$\Rightarrow$ Use sub-distribution of staphylococcus infection in the presence of the competing event (=discharge)

Go Back

Examples of models for informative censoring:

- F_T and F_C share common parameters:

$$T \sim N(\mu_T, \sigma) \quad \textit{and} \quad C \sim N(\mu_C, \sigma)$$

- Koziol-Green model:

$$1 - F_C(t) = [1 - F_T(t)]^{\gamma}$$

$\Rightarrow$ Dependence on the level of the distribution functions

Go Back