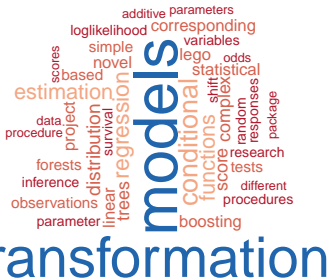# Smooth Transformation Models for Dependent Censoring

Sandra Siegfried and Torsten Hothorn

ICSDS 2024, Nice

Motivation

Independent censoring is a common assumption in survival analysis.

What to do if survival time $T \in \mathbb{R}^+$ and censoring time $C \in \mathbb{R}^+$ are dependent?

A Nonidentifiability Aspect of the Problem of Competing Risks

(crude survival probabilities/net survival probabilities)

ANASTASIOS TSIATIS*

Motivation

What to do if survival time $T \in \mathbb{R}^+$ and censoring time $C \in \mathbb{R}^+$ are dependent?

Copula Based Cox Proportional Hazards Models for Dependent Censoring

Negera Wakgari Deresa and Ingrid Van Keilegom

$$\mathbb{P}(T \leq t, C \leq c \mid \boldsymbol{X} = \boldsymbol{x}) =$$
$$\Phi_{0, \boldsymbol{R}(\lambda)} \left\{ \Phi^{-1} \left[\mathbb{P}(T \leq t \mid \boldsymbol{X} = \boldsymbol{x}) \right], \Phi^{-1} \left[\mathbb{P}(C \leq c \mid \boldsymbol{X} = \boldsymbol{x}) \right] \right\}$$

Follicular Cell Lymphoma Data

- $N = 541$ patients with early disease stage (I or II) and treated with Radiation Alone ("RT") or with Radiation and Chemotherapy ("CMT + RT").
- Time to relapse or death in remission.

	RT	CMTRT
Administrative censoring	131	62
Relapse	226	46
Death	66	10

Follicular Cell Lymphoma Competing Risks

10.32614/CRAN.package.SemiPar.depCens
10.32614/CRAN.package.tram

		SemiPar.depCens	tram
Disease relapse	chCMTRT	−0.352 (0.162)	−0.300 (0.154)
	age	0.418 (0.116)	0.435 (0.065)
	hgb	0.040 (0.069)	0.033 (0.058)
	clinstgI	−0.643 (0.127)	−0.575 (0.122)
Death in remission	(Intercept)	2.241 (1.058)	3.052
	chCMTRT	0.296 (0.303)	0.292
	age	−0.663 (0.112)	−0.829
	hgb	−0.030 (0.109)	−0.042
	clinstgI	0.635 (0.395)	0.684
	σ	0.955 (0.147)	0.871
Dependence	τ	0.769 (0.809)	0.774
	λ		−2.693 (1.089)
Computation Time		312.973 (50 Cores)	17.391 (1 Core)

World of Transformations

We want to invite you to

- understand (Gaussian) copula models for dependent censoring as nonparanormal models featuring marginal transformation models,
- stop the mistreatment of baseline cumulative hazards as "nuisance",
- accept smooth and parsimonously parameterised baseline cumulative hazards and other transformation functions,
- enjoy full likelihood estimation and inference,
- transfer core ideas to other domains (> 2 competing events, non-proportional hazards, prognostic biomarkers, etc).

Smooth Transformation Models for Dependent Censoring

1) Smooth Parametrisation

$$\mathbb{P}(T \le t \mid \boldsymbol{X} = \boldsymbol{x}) \;=\; \text{cloglog}^{-1}\left(h(t \mid \boldsymbol{\vartheta}) - \boldsymbol{x}^\top \boldsymbol{\beta}\right),$$

$$\mathbb{P}(C \le c \mid \boldsymbol{X} = \boldsymbol{x}) \;=\; \text{cloglog}^{-1}\left(\vartheta_1 + \vartheta_2 \log(c) - \boldsymbol{x}^\top \boldsymbol{\beta}_C\right).$$

with polynomial $h(t \mid \boldsymbol{\vartheta})$ in Bernstein form
(e.g. 10.1111/sjos.12291)

2) Parametrisation of correlation matrix as

$$\boldsymbol{R}(\lambda) \;=\; \begin{pmatrix} 1 & -\lambda/\sqrt{1+\lambda^2} \\ -\lambda/\sqrt{1+\lambda^2} & 1 \end{pmatrix}, \qquad \lambda \in (-\infty, \infty)$$

Smooth Transformation Models for Dependent Censoring

3) Formulate the joint distribution by a nonparanormal model

$$\mathbb{P}(T \leq t, C \leq c \mid \boldsymbol{X} = \boldsymbol{x}) =$$
$$\Phi_{0,\boldsymbol{R}(\lambda)} \left\{ \Phi^{-1} \left[\mathbb{P}(T \leq t \mid \boldsymbol{X} = \boldsymbol{x}) \right], \Phi^{-1} \left[\mathbb{P}(C \leq c \mid \boldsymbol{X} = \boldsymbol{x}) \right] \right\},$$

4) Evaluate likelihood

$$
\begin{aligned}
\text{Event at} \quad t: \quad & f_T(t|C > t, \boldsymbol{X} = \boldsymbol{x}) \times \mathbb{P}(C > t \mid \boldsymbol{X} = \boldsymbol{x}) \\
\text{Censored at} \quad t: \quad & f_C(t|T > t, \boldsymbol{X} = \boldsymbol{x}) \times \mathbb{P}(T > t \mid \boldsymbol{X} = \boldsymbol{x}) \\
\text{End of study at} \quad t: \quad & \mathbb{P}(T > t \wedge C > t \mid \boldsymbol{X} = \boldsymbol{x})
\end{aligned}
$$

5) Optimise jointly wrt ϑ, ϑ_1, ϑ_2, $\boldsymbol{\beta}$, $\boldsymbol{\beta}_C$, λ and perform standard ML inference

Cox Regression under Dependent Censoring

```r
library("tram")
...
Coxph(   ### marginal Cox + Weibull, Gaussian copula
    Surv(time, event = status) ~ ### censored, relapse,
                                 ### or death
    ch + age + hgb + clinstg,    ### covariates
    data = follic)
```
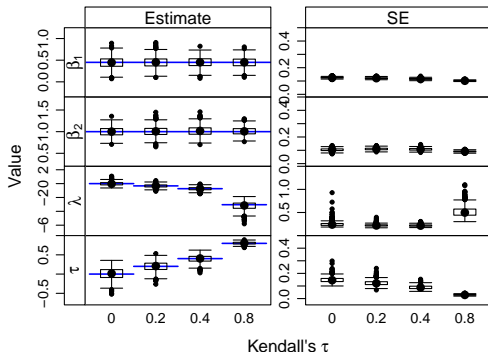
Coefficient	Estimate	Std. Error	95%-Wald CI	
β_{chCMTRT}	-0.300	0.154	-0.602 to	0.002
β_{age}	0.435	0.065	0.307 to	0.563
β_{hgb}	0.033	0.058	-0.080 to	0.146
β_{clinstgI}	-0.575	0.122	-0.814 to	-0.336
Log-Likelihood				
$-1'346.61$				

arXiv.2402.06428

Simulation Complex Baseline EoI

Deresa & Van Keilegom (2024) simulation, but with non-Weibull time-to-event of interest margin:

$$\mathbb{P}(T \leq t \mid \boldsymbol{X} = \boldsymbol{x}) = \text{cloglog}^{-1}\left(\text{cloglog}(F_{\chi_3^2}(t)) + x_1\beta_1 + x_2\beta_2\right)$$

Extension I: More than 2 Competing Events

Observe only one of $T_1, \ldots, T_J$, likelihood for $T_j = t$

$$f_j(t \mid \boldsymbol{X} = \boldsymbol{x}) \times \mathbb{P}\left(\bigcap_{\jmath \neq j} T_\jmath > t \mid T_j = t, \boldsymbol{X} = \boldsymbol{x}\right)$$

Marginal Cox model for one event of interest; Weibull for the remaining events

$$
\begin{aligned}
\mathbb{P}(T_1 \leq t \mid \boldsymbol{X} = \boldsymbol{x}) &= \text{cloglog}^{-1}\left(h(t \mid \boldsymbol{\vartheta}) - \boldsymbol{x}^\top \boldsymbol{\beta}_1\right), \\
\mathbb{P}(T_j \leq t \mid \boldsymbol{X} = \boldsymbol{x}) &= \text{cloglog}^{-1}\left(\vartheta_{j1} + \vartheta_{j2} \log(t) - \boldsymbol{x}^\top \boldsymbol{\beta}_j\right) \\
& \qquad j = 2, \ldots, J
\end{aligned}
$$
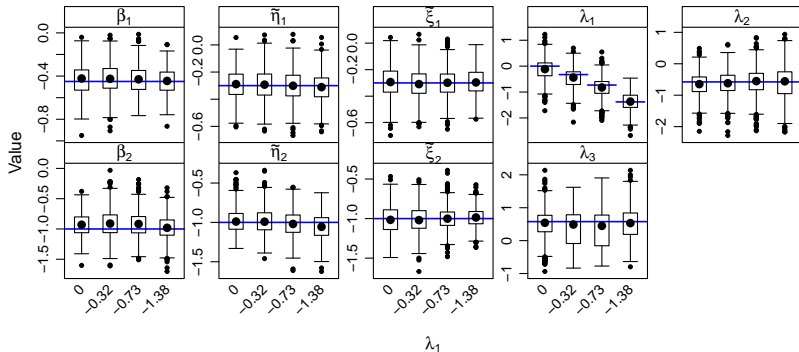
Extension I: More than 2 Competing Events

Parameterise correlation by scaled inverse Cholesky factor

$$
\begin{aligned}
\boldsymbol{R}(\boldsymbol{\lambda}) &= \boldsymbol{\Omega}(\boldsymbol{\lambda})^{-1}\boldsymbol{\Omega}(\boldsymbol{\lambda})^{-\top} \\
\boldsymbol{\Omega}(\boldsymbol{\lambda}) &= \boldsymbol{\Lambda}(\boldsymbol{\lambda})\operatorname{diag}(\boldsymbol{\Lambda}^{-1}(\boldsymbol{\lambda})\boldsymbol{\Lambda}^{-\top}(\boldsymbol{\lambda}))^{1/2} \\
\boldsymbol{\Lambda} &= \boldsymbol{\Lambda}(\boldsymbol{\lambda}) = \begin{pmatrix}
1 & & & & \\
\lambda_{21} & 1 & & 0 & \\
\lambda_{31} & \lambda_{32} & 1 & & \\
\vdots & \vdots & & \ddots & \\
\lambda_{J1} & \lambda_{J2} & \dots & \lambda_{J,J-1} & 1
\end{pmatrix}
\end{aligned}
$$

featuring $\boldsymbol{\lambda} = (\lambda_{21}, \dots, \lambda_{J,J-1})^{\top} \in \mathbb{R}^{J(J-1)/2}$ *unconstraint* parameters (10.1111/sjos.12291).
Log-likelihood + scores: 10.32614/CRAN.package.mvtnorm

Simulation Competing Risk with $J = 3$

Extension II: Interval Censoring

We observe T_j in interval $(\underline{t}, \overline{t}]$, the likelihood is then

$$\mathbb{P}\left(\underline{t} < T_j \le \overline{t}, \bigcap_{\jmath \neq j}^{J} T_\jmath > \overline{t} \mid \boldsymbol{X} = \boldsymbol{x}\right)$$

Smooth parameterisation of baseline log-cumulative hazard $h(t \mid \vartheta)$ and application of Genz' algorithm allows evaluation and optimisation of corresponding log-likelihood

Extension III: Non-proportional Hazards

Replace marginal Cox proportional hazards model

$$\mathbb{P}(T \leq t \mid \boldsymbol{X} = \boldsymbol{x}) = \text{cloglog}^{-1}\left(h(t \mid \boldsymbol{\vartheta}) - \boldsymbol{x}^{\top}\boldsymbol{\beta}\right)$$

with a location-scale transformation model

$$\mathbb{P}(T \leq t \mid \boldsymbol{X} = \boldsymbol{x}) = \text{cloglog}^{-1}\left(h(t \mid \boldsymbol{\vartheta})\exp(\boldsymbol{x}^{\top}\boldsymbol{\gamma}) - \boldsymbol{x}^{\top}\boldsymbol{\beta}\right)$$

allowing crossing hazards.

10.1080/00031305.2023.2203177

Extension IV: Prognostic Biomarkers

With $X \in \mathbb{R}$ denoting a prognostic biomarker, the model

$$\mathbb{P}(T \leq t, C \leq c, X \leq x) =$$
$$\Phi_{0,\boldsymbol{R}(\boldsymbol{\lambda})} \left\{ \Phi^{-1}\left[\mathbb{P}(T \leq t)\right], \Phi^{-1}\left[\mathbb{P}(C \leq c)\right], \Phi^{-1}\left[\mathbb{P}(X \leq x)\right] \right\}$$

can be used to derive and evaluate time-dependent ROC curves under dependent censoring.

Estimation and Inference

- Negative log-likelihood in general non-convex
- In many cases relevant here, it is bi-convex in the marginal vs. λ parameters
- Nonparanormal models contain all special cases discussed in this presentation

arXiv.2408.17346

- Continuous, discrete, and mixed Gaussian copula log-likelihood (`mvtnorm::ldpmvnorm()`) and score (`mvtnorm::sldpmvnorm()`) functions available
- Nonparanormal model estimation and ML inference implemented in `tram::Mmlt()`
- `tram::Coxph()` and `tram::Compris()` user interfaces

10.32614/CRAN.package.mvtnorm,
10.32614/CRAN.package.tram

https://ctm.R-forge.R-project.org/