2019 6/3-6/7; MMR2019
11th Conference on Mathematical Methods in Reliability, Hong Kong

A Frailty-Copula Model for Dependent Competing Risks in Reliability Theory

Takeshi Emura
National Central University

Joint work with Yin-Chen Wang

June 6, 2019

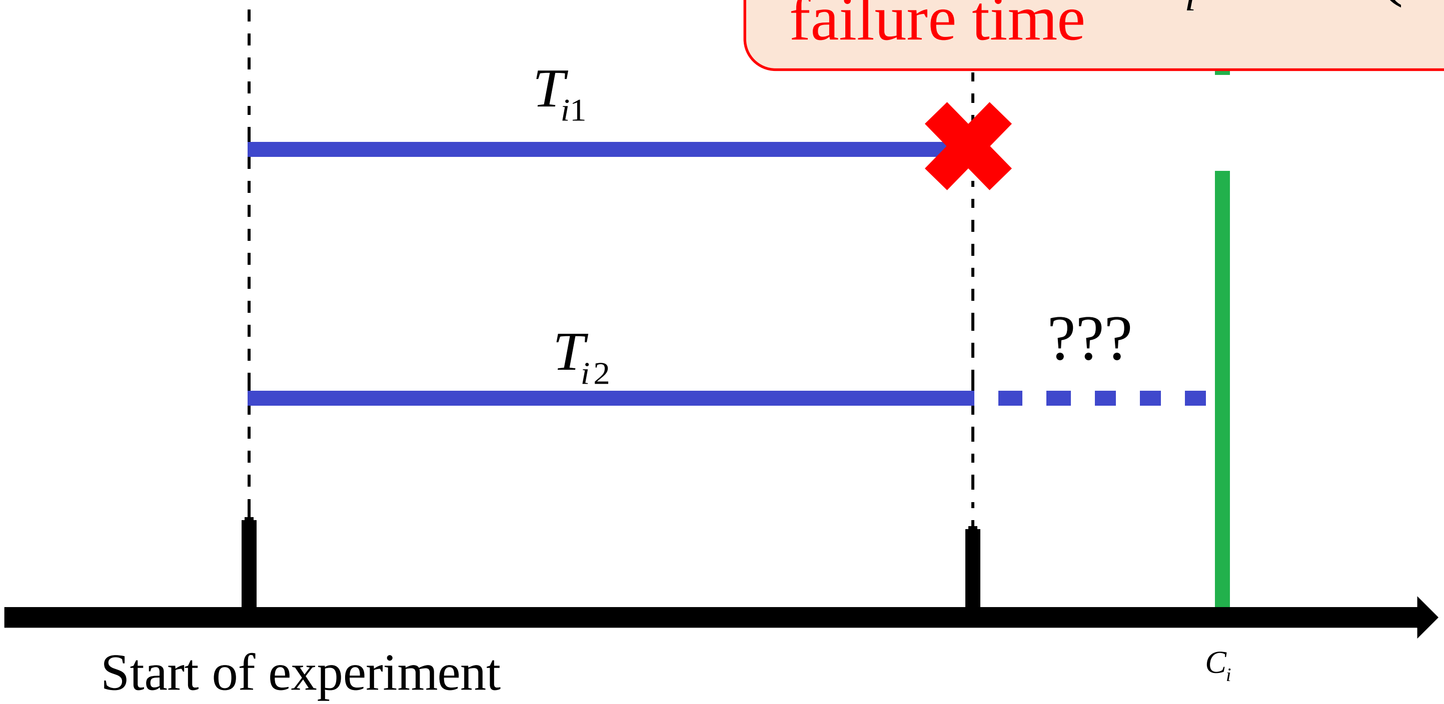
Bivariate competing risks

T_{i1} : Failure time of Event 1

T_{i2} : Failure time of Event 2

C_i : Censoring time

Observed failure time $T_i = \min(T_{i1}, T_{i2}, C_i)$



Bivariate competing risks models

- Bivariate Weibull models (Moeschberger 1974; David and Moeschberger1978)
- Bivariate Marshall-Olkin Weibull model (Fan et al. 2019)
- Bivariate Pareto model (Sankaran and Kundu 2014; Shih et al. 2019)
- Bivariate normal model

  (Nádas 1971; Moeschberger 1974; Basu and Ghosh 1978)
- Independent Lindley model (Mazucheli and Achcar 2011)
- Bivariate frailty model (Liu 2012; Lo et al. 2017)
- Bivariate copula models (Hsu et al. 2017; Emura et al. 2017

  Emura and Michimae 2017; Zhou et al. 2018)

We propose a hybrid (frailty + copula) model

Frailty model (Liu 2012, Technometrics)

- Conditional hazard function

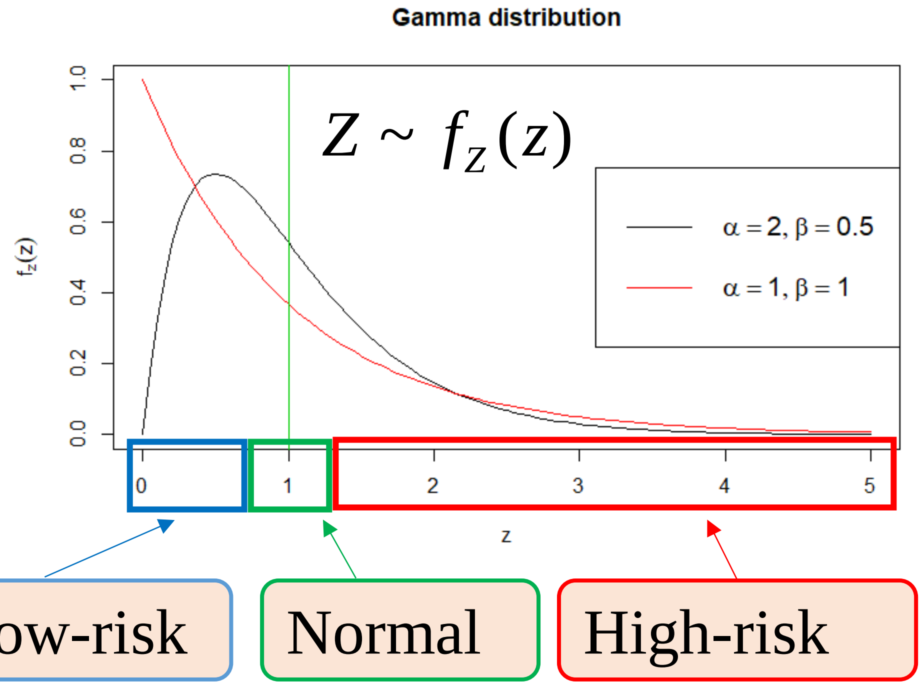
$$h_{T_j|Z}(t \mid z) = z \times h_0(t)$$

Baseline ($Z = 1$)

-Heterogeneity parameter:

$$\eta = Var(Z)$$

$$f_Z(z) = \frac{1}{\Gamma(1/\eta)\eta^{1/\eta}} z^{1/\eta - 1} \exp\left(-\frac{z}{\eta}\right)$$



$$E[Z] = 1$$

Frailty model (Liu 2012)

-Conditional independence given *z*

$$\Pr(T_1 > t_1, T_2 > t_2 \mid z) = \prod_{i=1}^{2} \Pr(T_i > t_i \mid z)$$

- Marginal Weibull model

$$\Pr(T_j > t \mid z) = S_{T_j|Z}(t \mid z) = \exp\left[-z \left\{ \frac{t}{\exp(\mu_j)} \right\}^{\frac{1}{\sigma_j}} \right]$$

- Integrating out unobserved *z*

$$S_{T_{i1},T_{i2}}(t_1, t_2) = \int_0^\infty S_{T_{i1},T_{i2}|Z_i}(t_1, t_2 \mid z) f_Z(z) dz = \left(1 + \eta \sum_{j=1}^{2} \left\{ \frac{t_j}{\exp(\mu_j)} \right\}^{\frac{1}{\sigma_j}} \right)^{-\frac{1}{\eta}}$$

$$\tau_\eta = \frac{2}{\eta + 2}: \quad \text{Kendall's tau}$$

Motivated by a joint frailty-copula model (Emura et al. 2017):

we relax the conditional independence by a copula

$$S_{T_1,T_2}(t_1, t_2 \mid z) = C_\theta\left[\, S_{T_1}(t_1 \mid z), S_{T_2}(t_2 \mid z) \,\right]$$

Dependence parameter

Gumbel copula

$$C_\theta(u, v) = \exp[-\{(-\log u)^{\theta+1} + (-\log v)^{\theta+1}\}^{1/(\theta+1)}]$$

$$\theta = 0$$

Independence copula $C_\theta(u, v) = uv$

Proposed model

Gamma-Gumbel model

$$\Pr(T_{i1} > t_1, T_{i2} > t_2 \mid z) = S_{T_{i1}, T_{i2} \mid Z_i}(t_1, t_2 \mid z)$$

$$= \exp\left(-z\left[\left\{ \frac{t_1}{\exp(\mu_1)} \right\}^{(\theta+1)/\sigma_1} + \left\{ \frac{t_2}{\exp(\mu_2)} \right\}^{(\theta+1)/\sigma_2} \right]^{1/(\theta+1)} \right)$$

$\theta = 0$

Reduce to Liu (2012)'s model

$$S_{T_{i1}, T_{i2} \mid Z_i}(t_1, t_2 \mid z)$$

$$= \exp\left[-z\left\{ \frac{t_1}{\exp(\mu_1)} \right\}^{\frac{1}{\sigma_1}} \right] \exp\left[-z\left\{ \frac{t_2}{\exp(\mu_2)} \right\}^{\frac{1}{\sigma_2}} \right].$$

Integrating out unobserved frailty

$$S_{T_{i1},T_{i2}}(t_1, t_2) = \int_0^\infty S_{T_{i1},T_{i2}|Z_i}(t_1, t_2|z)\, f_Z(z)dz$$

$$= \left(1 + \eta \left[\sum_{j=1}^{2} \left\{ \frac{t_j}{\exp(\mu_j)} \right\}^{(\theta+1)/\sigma_j} \right]^{1/(\theta+1)} \right)^{-\frac{1}{\eta}}$$

Parameters: $(\mu_1, \mu_2, \sigma_1, \sigma_2, \eta, \theta)$

$(\mu_1, \mu_2, \sigma_1, \sigma_2, \eta)$

$\theta = \text{Fixed}$

Identifiability issue

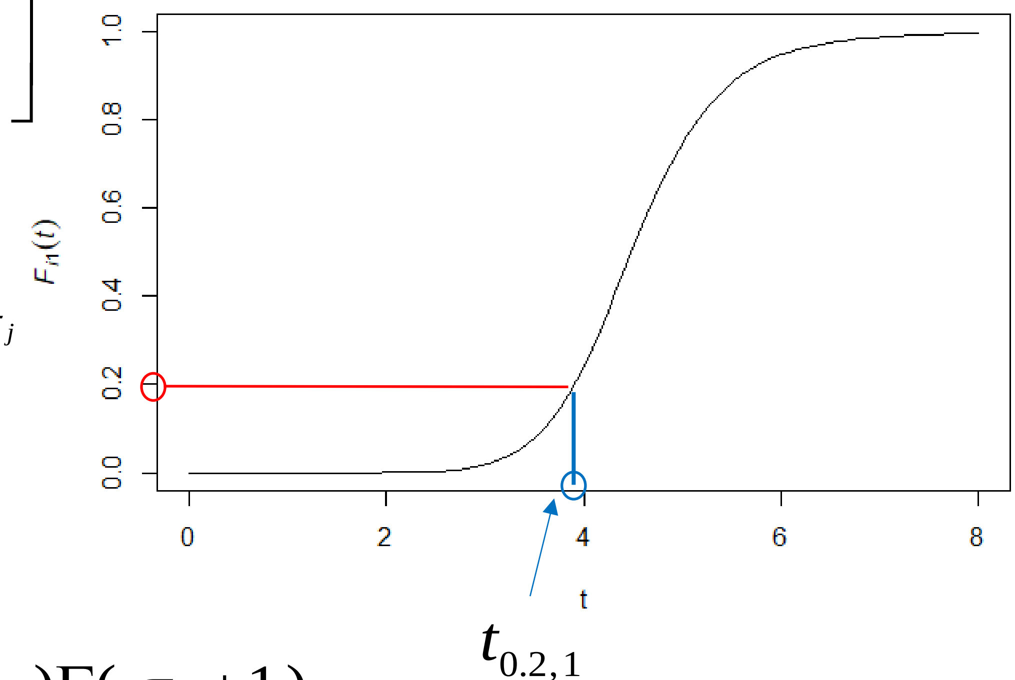
Proposed model: properties

- Three-parameter Burr XII model (Burr 1942).

$$S_{T_{ij}}(t) = \left[1 + \eta \left\{ \frac{t}{\exp(\mu_j)} \right\}^{1/\sigma_j} \right]^{-1/\eta}$$

- The *p*-quantile:

$$t_{p,j} = \exp(\mu_j) \left\{ \frac{(1-p)^{-\eta} - 1}{\eta} \right\}^{\sigma_j}$$



- Mean & Variance exists

$$E(T_{ij}) = \frac{\exp(\mu_j)\Gamma(1/\eta - \sigma_j)\Gamma(\sigma_j + 1)}{\eta^{\sigma_j + 1}\Gamma(1/\eta + 1)}$$

Competing risks data

T_{i1}: failure time due to cause 1

T_{i2}: failure time due to cause 2

C_i : independent censoring time

$T_i = \min(T_{i1}, T_{i2}, C_i)$: observed failure time

$\delta_i = \mathbf{I}(T_{i1} < T_{i2}, T_{i1} < C_i)$: indicator of failure cause 1

$\delta_i^* = \mathbf{I}(T_{i2} < T_{i1}, T_{i2} < C_i)$: indicator of failure cause 2

Our data is $(T_i, \delta_i, \delta_i^*)$, for $i = 1, 2, \cdots, n.$

Maximum likelihood inference

- The log-likelihood function

$$\ell(\mu_1, \mu_2, \sigma_1, \sigma_2, \eta, \theta) = \sum_{i=1}^{n} \delta_i \log f(T_i, 1) + \sum_{i=1}^{n} \delta_i^* \log f(T_i, 2) + \sum_{i=1}^{n} (1 - \delta_i - \delta_i^*) \log S_{T_{i1}, T_{i2}}(T_i, T_i).$$

- Sub-density

$$f(t, j) = \left(1 + \eta \left[\left\{ \frac{t}{\exp(\mu_1)} \right\}^{\frac{\theta+1}{\sigma_1}} + \left\{ \frac{t}{\exp(\mu_2)} \right\}^{\frac{\theta+1}{\sigma_2}} \right]^{\frac{1}{\theta+1}} \right)^{-\frac{1+\eta}{\eta}} \times \left[\left\{ \frac{t}{\exp(\mu_1)} \right\}^{\frac{\theta+1}{\sigma_1}} + \left\{ \frac{t}{\exp(\mu_2)} \right\}^{\frac{\theta+1}{\sigma_2}} \right]^{\frac{1}{\theta+1} - 1} \times \frac{t^{\frac{\theta+1}{\sigma_j} - 1}}{\sigma_j \exp\left\{ \mu_j \left(\frac{\theta+1}{\sigma_j} \right) \right\}}.$$

Goodness-of-fit (Parametric vs. Nonpar)

- Sub-distribution functions

$$F(t, j) \equiv \int_0^t f(u, j)\,du$$

- Parametric estimation of sub-distribution function

$$F_{\hat{\Phi}}(t, j) = \int_0^t \left(1 + \hat{\eta} \left[\left\{ \frac{s}{\exp(\hat{\mu}_1)} \right\}^{\frac{\theta+1}{\hat{\sigma}_1}} + \left\{ \frac{s}{\exp(\hat{\mu}_2)} \right\}^{\frac{\theta+1}{\hat{\sigma}_2}} \right]^{\frac{1}{\theta+1}} \right)^{-\frac{1+\hat{\eta}}{\hat{\eta}}}$$

$$\times \left[\left\{ \frac{s}{\exp(\hat{\mu}_1)} \right\}^{\frac{\theta+1}{\hat{\sigma}_1}} + \left\{ \frac{s}{\exp(\hat{\mu}_2)} \right\}^{\frac{\theta+1}{\hat{\sigma}_2}} \right]^{\frac{1}{\theta+1}-1} \frac{s^{\frac{\theta+1}{\hat{\sigma}_j}-1}}{\hat{\sigma}_j \exp\left\{ \hat{\mu}_j \left(\frac{\theta+1}{\hat{\sigma}_j} \right) \right\}} ds.$$

Goodness-of-fit(Parametric vs. Nonpar)

- Nonparametric estimation

$$\widehat{F}(t, j) = \sum_{i:T_{(i)} \le t} \widehat{S}(t)\frac{d_{ij}}{n_i}, \quad j = 1, 2$$

where

$$n_i = \sum_{j=1}^{n} \mathbf{I}\left(T_j \ge T_{(i)} \right)$$

$$d_i = \sum_{j=1}^{n} \mathbf{I}\left(T_j = T_{(i)} \right)$$

$$d_{i1} = \sum_{j=1}^{n} \delta_i \mathbf{I}\left(T_j = T_{(i)} \right)$$

$$d_{i2} = \sum_{j=1}^{n} \delta_i^* \mathbf{I}\left(T_j = T_{(i)} \right)$$

$$\widehat{S}(t) = \prod_{i:T_{(i)} \le t} (1 - d_i / n_i).$$

- The Cramér-von Mises statistic

$$CvM = \| \text{ Parametric - Nonparametric } \|^2$$

$$= \sum_{i=1}^{n} \delta_i \{ F_{\widehat{\Phi}}(T_i, 1) - \widehat{F}(T_i, 1) \}^2 + \sum_{i=1}^{n} \delta_i^* \{ F_{\widehat{\Phi}}(T_i, 2) - \widehat{F}(T_i, 2) \}^2.$$

Simulation design

- Generate samples $\{ (T_{i1}, T_{i2}), \quad i = 1, 2, \ldots, n \}$

  Step 1: Generate $Z \sim Gamma(1/\eta, \eta)$.

  Step 2: Generate $(U_1, U_2) \sim C_\theta(u,v), \quad \theta = 6$

  Step 3: Set $T_1 = S_{01}^{-1}(U_1^{1/Z})$ and $T_2 = S_{02}^{-1}(U_2^{1/Z})$.

- Independent censoring $C_i \sim U(0, w), \quad i = 1, 2, \ldots, n.$

  $w = 14$ (20% censoring)

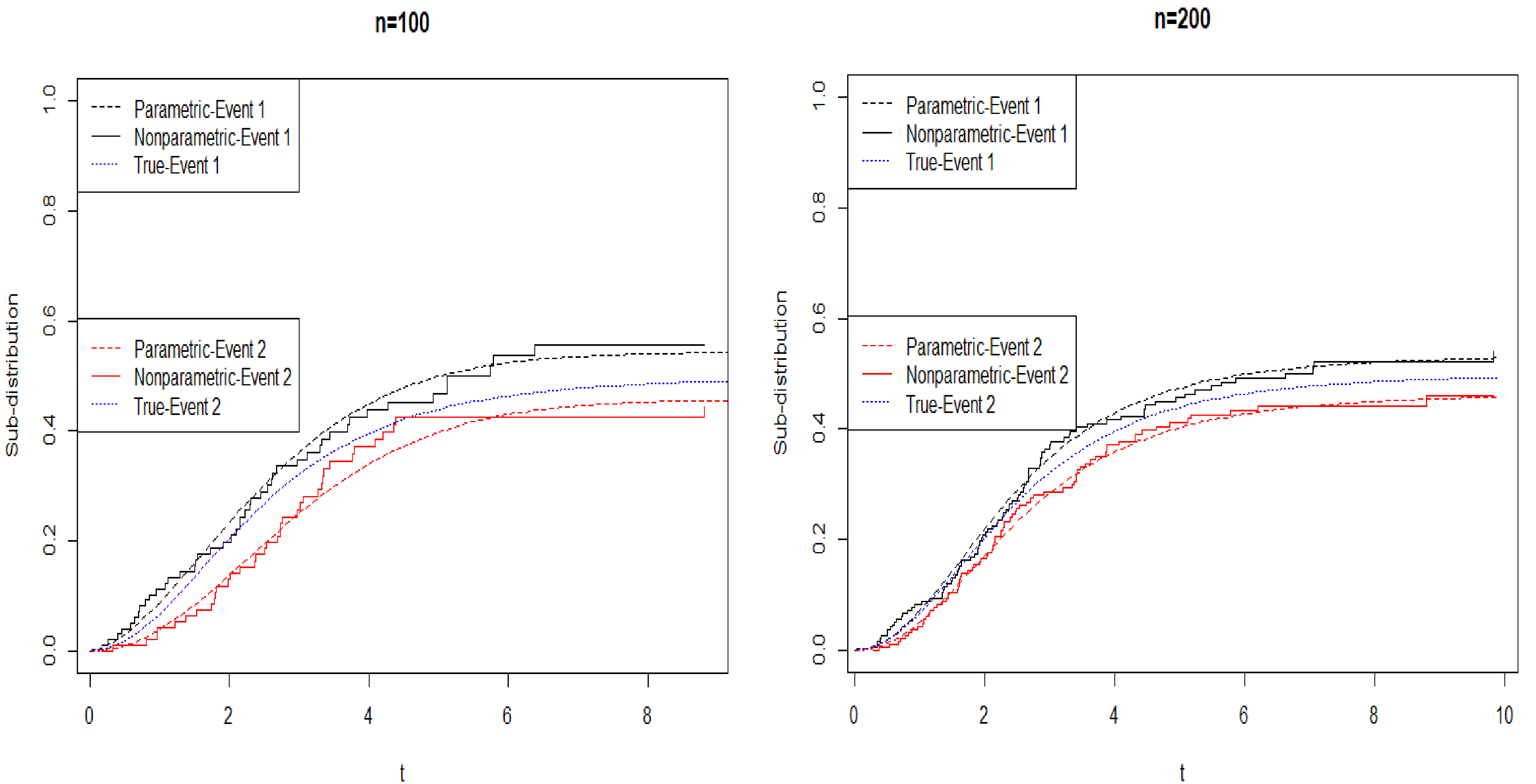
- Competing risks data

$$T_i = \min(T_{i1}, T_{i2}, C_i), \; \delta_i = \mathbf{I}(T_i = T_{i1}), \; \delta_i^* = \mathbf{I}(T_i = T_{i2}).$$

Simulation results

	n	Parameter	True	Mean	SD	SE	CP%	ERR%
CEN = 20% T1 = 40% T2 = 40%	100	μ_1	1.0000	1.0108	0.1376	0.1269	0.9091	0.01
		μ_2	1.0000	1.0107	0.1371	0.1269	0.9131	0.01
		σ_1	0.5000	0.5031	0.0816	0.0771	0.9313	0.01
		σ_2	0.5000	0.5034	0.0815	0.0772	0.9374	0.01
		η	0.5000	0.5068	0.3731	0.3435	0.9576	0.01
		$t_{0.01,1}$	0.2729	0.2846	0.0874	0.0789	0.9192	0.01
		$t_{0.1,1}$	0.8941	0.9054	0.1261	0.1179	0.9273	0.01
		$t_{0.5,1}$	2.4741	2.4933	0.2148	0.2034	0.9354	0.01
	200	μ_1	1.0000	0.9890	0.1873	0.0909	0.9379	0.002
		μ_2	1.0000	0.9893	0.1789	0.0909	0.9419	0.002
		σ_1	0.5000	0.4973	0.0572	0.0550	0.9379	0.002
		σ_2	0.5000	0.4972	0.0574	0.0550	0.9399	0.002
		η	0.5000	0.5220	0.2573	0.2456	0.9519	0.002
		$t_{0.01,1}$	0.2729	0.2796	0.0577	0.0561	0.9439	0.002
		$t_{0.1,1}$	0.8941	0.8969	0.0911	0.0832	0.9539	0.002
		$t_{0.5,1}$	2.4741	2.4664	0.1810	0.1433	0.9539	0.002

Results (Nonparametric vs. Parametric)



ARC-1 VHF radio data (Mendenhall and Hader 1958)

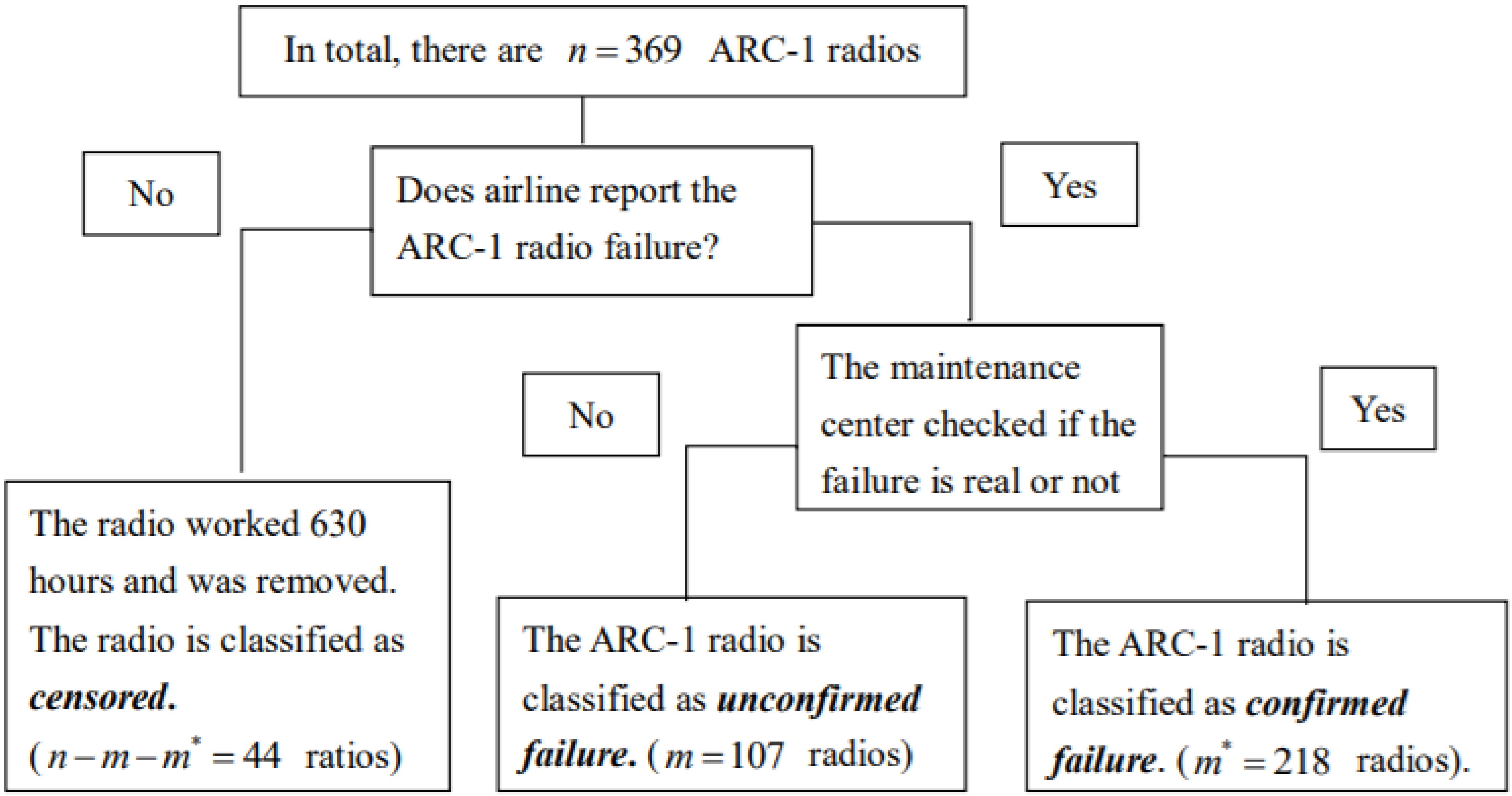


Figure 11: The ARC-1 VHF radio dataset from Mendenhall and Hader (1958)

Data analysis

- The summary of ARC-1 VHF radio data

	Confirmed $\delta_i = 1$	Unconfirmed $\delta_i^* = 1$	Censored $\delta_i = \delta_i^* = 0$
The number of events (event rate %)	218(59%)	107(29%)	44(12%)
Average (hours)	$\bar{T}_1 = 229.61$	$\bar{T}_2 = 191.20$	$C = 630$ (fixed)

- Fit the model

$$\Pr(T_{i1} > t_1, T_{i2} > t_2 \mid z) = S_{T_{i1}, T_{i2} \mid Z_i}(t_1, t_2 \mid z)$$

$$= \exp\left(-z \left[\left\{ \frac{t_1}{\exp(\mu_1)} \right\}^{(\theta+1)/\sigma_1} + \left\{ \frac{t_2}{\exp(\mu_2)} \right\}^{(\theta+1)/\sigma_2} \right]^{1/(\theta+1)} \right)$$

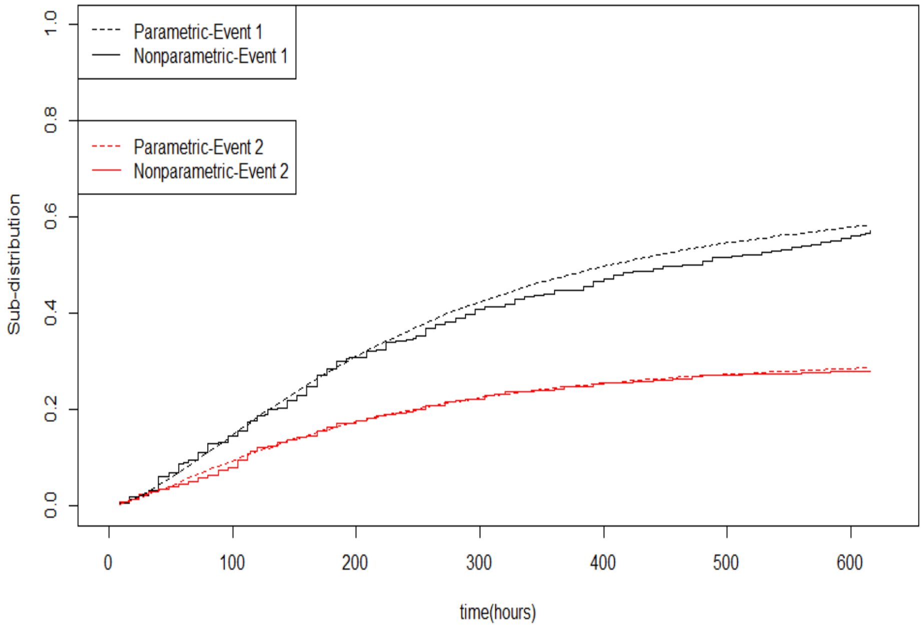
given $\theta = 0,\ 0.33,\ 1,$ or 3

Results of Data Aanlysis

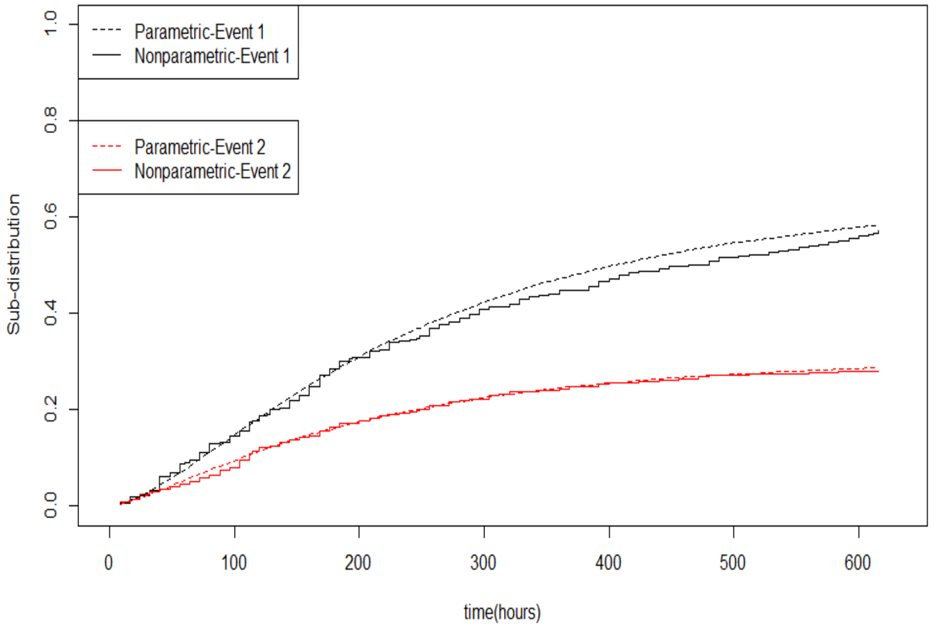
Table 3: The fitted results of the gamma-Gumbel copula model to the ARC-1 data.

	Parameter estimate (95% CI)			
	$\theta = 0$ ($\tau_\theta^{\|Z} = 0$)	$\theta = 0.33$ ($\tau_\theta^{\|Z} = 0.25$)	$\theta = 1$ ($\tau_\theta^{\|Z} = 0.5$)	$\theta = 3$ ($\tau_\theta^{\|Z} = 0.75$)
μ_1	5.70 (5.46-5.95)	5.63 (5.40-5.86)	5.56 (5.34-5.78)	5.50 (5.29-5.71)
μ_2	6.24 (5.88-6.61)	6.04 (5.72-6.36)	5.84 (5.56-6.12)	5.64 (5.40-5.88)
σ_1	0.65 (0.54-0.79)	0.66 (0.56-0.80)	0.67 (0.57-0.80)	0.69 (0.58-0.82)
σ_2	0.76 (0.62-0.95)	0.75 (0.61-0.91)	0.73 (0.60-0.88)	0.71 (0.59-0.86)
η	0.65 (0.31-1.34)	0.64 (0.31-1.33)	0.64 (0.31-1.32)	0.62 (0.30-1.32)
$t_{0.01,1}$	14.75 (9.78-22.23)	13.16 (8.81-19.69)	11.75 (7.80-17.59)	10.45 (6.87-15.92)

Goodness-of-fit (Parametric v.s Nonpara)



(a) $\theta = 0 \,(\, \tau_{\theta}^{|Z} = 0 \,)$



(b) $\theta = 3 \,(\, \tau_{\theta}^{|Z} = 0.75 \,)$

Future works

- P-value of goodness-of-fit

  - Parametric bootstrap

- Semipar. estimation (e.g. spline) of the sub-distribution function

  - Parametric vs. Nonparametric vs. Semiparametric

  Goodness-of-fit for a copula

- Covariate

  - AFT models (location-scale regression)

- Selection of copula parameter

  - Profile likelihood does not work well

  - Minimum distance CvM statistics

References

- Basu AP, Ghosh JK (1978) Identifiability of the multinormal and other distributions under competing risks model. Journal of Multivariate Analysis 8, 413-429.
- Burr IW (1942) Cumulative frequency functions. Ann Math Stat 13(2), 215-232.
- David HA, Moeschberger ML (1978) The Theory of Competing Risks. London: Griffin.
- Emura T, Nakatochi M, Murotani K, Rondeau V (2017) A joint frailty-copula model between tumour progression and death for meta-analysis, Stat Methods Med Res 26: 2649-66.
- Fan TH, Wang YF, Ju SK (2019). A competing risks model with multiply censored reliability data under multivariate Weibull distributions. IEEE Transactions on Reliability.
- Liu X (2012) Planning of accelerated life tests with dependent failure modes based on a gamma frailty model. Technometrics 54(4), 398-409.
- Mendenhall W, Hader RJ (1958) Estimation of parameters of mixed exponential distributed failure time distribution from censored life test data. Biometrika 45, 504-520.
- Moeschberger ML (1974) Life tests under dependent competing causes of failure. Technometrics 16(1), 39-47.
- Nádas A (1971) The distribution of the identified minimum of a normal pair determines' the distribution of the pair. Technometrics 13(1), 201-202.
- Nelsen RB (2006) An Introduction to Copulas, (2nd ed.). Springer, New York.