Universität Rostock

Traditio et Innovatio

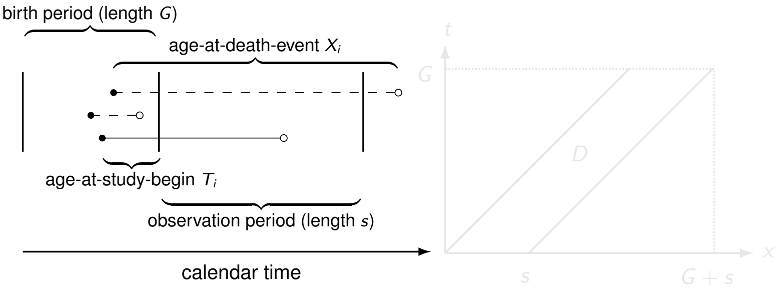
Gumbel-type copulas as model for truncation dependence

Rafael Weissbach & Anne-Marie Toparkus
University of Rostock, Chair of Statistics and Econometrics

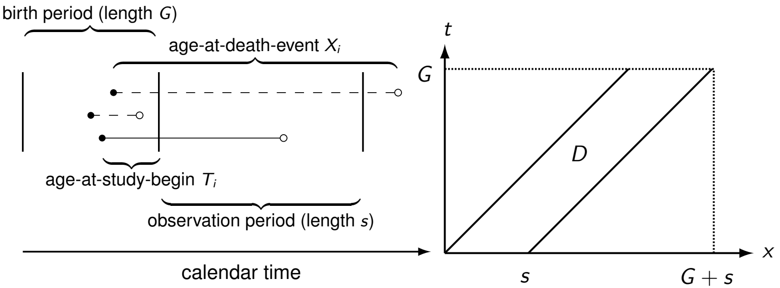
Table of contents

Table of contents

1 Introduction



1 Introduction



Simplifying assumptions:

- Independence of age at truncation T and duration X
  - – conflict with progress of life expectancy
  - ↪ copula dependence model and parametric test on the copula parameter
- Exponentially distributed X, uniformly distributed T
  - – possibly overly simple form
  - ↪ create a bivariate Kolmogorov-Smirnov type test

Table of contents

2 Population model and Likelihood

Assumptions:

(A1) Marginal distributions: $T_i \sim Uni[0, G]$ and $X_i \sim Exp(\theta)$ with $\theta \in [\varepsilon_\theta, 1/\varepsilon_\theta]$

(A2) Gumbel-Barnett copula with $\vartheta \in [0, 1 - \varepsilon_\vartheta]$:

$$C_\vartheta(u, v) = u + v - 1 + (1 - u)(1 - v)e^{-\vartheta \log(1-u) \log(1-v)}$$

- Bivariate exponential distribution after Gumbel (1960)

$$F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y-\vartheta xy}$$

- $(X_i, T_i)^T : (\Omega, \mathcal{A}, P_\theta) \to (S, \mathcal{B})$, $i = 1, \ldots, n$, $n \in \mathbb{N}$, u.i.v.
- f_θ joint density of $(X_i, T_i)^T$ wrt. P_θ for $x > 0$ and $0 < t < G$

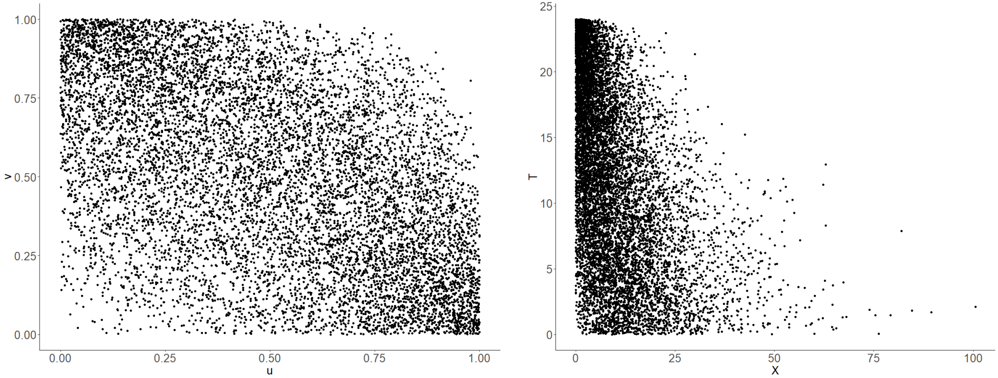


Abbildung: Scatter plot of $10,000$ bivariate random draws using Gumbel-Barnett copula ($\vartheta = 1$, $\tau = -0{,}36$) left: uniform margins; right: uniform and exponential margins of (A1)

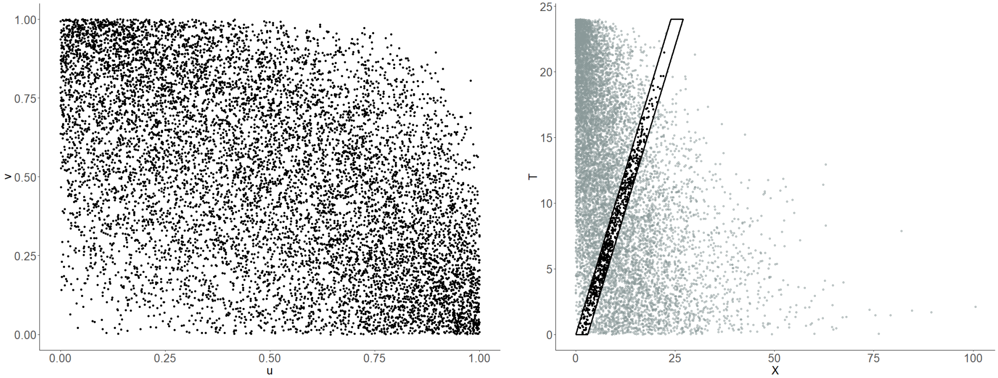


Abbildung: Scatter plot of 10,000 bivariate random draws using Gumbel-Barnett copula ($\vartheta = 1$, $\tau = -0{,}36$) left: uniform margins; right: uniform and exponential margins of (A1)

Truncation:

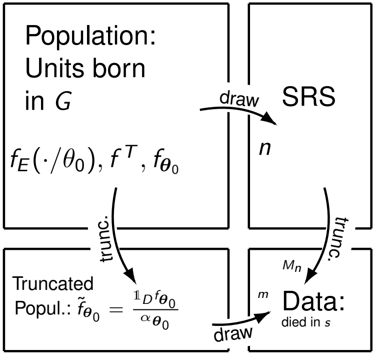
- Observation parallelogram

(A3) $D := \{(x,t)^T | 0 < t \leq x \leq t+s, t \leq G\}$

- Notation of observation by $(\widetilde{X}_j, \widetilde{T}_j)^T$, $j = 1, \ldots, M_n \leq n$, with $M_n = \sum_{i=1}^n \mathbb{1}_{\{(X_i,T_i)^T \in D\}}$

- Selection probability $$\alpha_{\boldsymbol{\theta}} = P_{\boldsymbol{\theta}}(T_i \leq X_i \leq T_i + s) = \int_0^G \int_t^{t+s} f_{\boldsymbol{\theta}}(x,t)\, \mathrm{d}x\, \mathrm{d}t$$

Design


Population:
Units born
in G
$f_E(\cdot/\theta_0), f^T, f_{\boldsymbol{\theta}_0}$
draw
SRS
n
trunc.
trunc.
M_n
Truncated
Popul.: $\tilde{f}_{\boldsymbol{\theta}_0} = \frac{\mathbb{1}_D f_{\boldsymbol{\theta}_0}}{\alpha_{\boldsymbol{\theta}_0}}$
m
Data:
died in s
draw

Likelihood:

- Standard results for point processes (see e.g. Reiss, 1993)
- Loglikelihood

$$\log \ell(\boldsymbol{\theta}, n) \approx \sum_{j=1}^{M_n} \log n f_{\boldsymbol{\theta}}(\widetilde{X}_j, \widetilde{T}_j) + (G+s)G - n\alpha_{\boldsymbol{\theta}}$$

$$= \sum_{i=1}^{n} \mathbb{1}_{\{(X_i, T_i)^T \in D\}} \log n f_{\boldsymbol{\theta}}(X_i, T_i) + (G+s)G - n\alpha_{\boldsymbol{\theta}}$$

- Maximize by replacing n with $n = M_n/\alpha_{\boldsymbol{\theta}}$, yields observability and reduces dimension of score function

Table of contents

3.1 Asymptotic distribution

- Hypothesis of independence: $H_0 : \vartheta_0 = 0$

$\hookrightarrow$ distribution of test statistic $T_n = \sqrt{n}\hat{\vartheta}_n/\sigma_\vartheta$ under H_0?

Preliminaries:

$\Rightarrow$ Under (A1) – (A3) holds weak consistency for $\boldsymbol{\theta}_0 \in [\varepsilon_\theta, 1/\varepsilon_\theta] \times [0, 1 - \varepsilon_\vartheta]$:

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0 \text{ for } n \to \infty$$

$\Rightarrow$ Under (A1) – (A3) holds asymptotic normality for $\boldsymbol{\theta}_0 \in (\varepsilon_\theta, 1/\varepsilon_\theta) \times (0, 1 - \varepsilon_\vartheta)$:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1}\right)$$

Theorem 3.3 Asympt. distrib. for $\vartheta_0 = 0$ (Moran, 1971)

Under assumptions (A1)-(A3), $\boldsymbol{\theta}_0 \in (\varepsilon_\theta, 1/\varepsilon_\theta) \times \{0\}$, the sequence $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges for $n \to \infty$ in distribution towards the mixture of distributions

$$\Phi_{\boldsymbol{\theta}_0}(a_\theta, a_\vartheta) = \frac{1}{2} F_1^{\boldsymbol{\theta}_0}(a_\theta, a_\vartheta) + \frac{1}{2} F_2^{\boldsymbol{\theta}_0}(a_\theta, a_\vartheta),$$

where

- $F_1^{\boldsymbol{\theta}_0}$ is defined on $\mathbb{R} \times \mathbb{R}_{>0}$ and corrsponds to $2 \cdot \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$.
- $F_2^{\boldsymbol{\theta}_0}$ is the distribution of $\sigma^{(2)} \widetilde{Y}_1$ on $\mathbb{R} \times \{0\}$, where $\sigma^{(2)} := (\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}$.
- The distribution of $\widetilde{Y}_1$ is the distribution of Y_1 conditional on

$$Y_2 - \mathcal{I}(\boldsymbol{\theta}_0)_{12} \cdot \sigma^{(2)} Y_1 \leq 0, \quad \text{with } (Y_1, Y_2)^T \sim \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)).$$

3.2 Stationarity of enterprise lifespans

Positive trend:

- Consideration of $m_n = 55\,279$ German enterprise lifespans $\widetilde{X}_j$ ending (due to insolvency) between 2014 and 2016
- Foundation period between 1990 and 2013 $\Rightarrow$ $\widetilde{T}_j$ age at the beginning of 2014
- Results:

Copula	$\hat{\theta}_n$	$\hat{\vartheta}_n$	$1/\hat{\theta}_n$	τ	$\alpha_{\hat{\theta}_n}$	$\hat{n}$	$\hat{\sigma}_\vartheta/\sqrt{n}$
GB	0.083	0	12.11	0	0.095	579 089	$5.19 \cdot 10^{-3}$

- No rejection of independence hypothesis ($t_{crit} = 1.645$, $\alpha = 0.05$)

$\Rightarrow$ Increasing lifespan with later foundations can not be expected

Negative trend:

- Rotated Gumbel-Barnett copula with $\vartheta^{RGB} \in [0, 1 - \varepsilon_\vartheta]$, Farlie-Gumbel-Morgenstern copula with $\vartheta^{FGM} \in [-1, 1]$

Copula	$\hat{\theta}_n$	$\hat{\vartheta}_n$	$1/\hat{\theta}_n$	τ	$\alpha_{\hat{\theta}_n}$
GB	0.083	0	12.11	0	0.095
RGB	0.083	0.06	12.05	0.029	0.097
FGM	0.082	0.10	12.24	0.023	0.098

Copula	$\hat{n}$	$\hat{\sigma}_\vartheta/\sqrt{n}$	T_n
GB	579 089	$5.19 \cdot 10^{-3}$	0
RGB	571 018	$5.82 \cdot 10^{-3}$	10.315
FGM	566 816	$1.81 \cdot 10^{-2}$	5.709

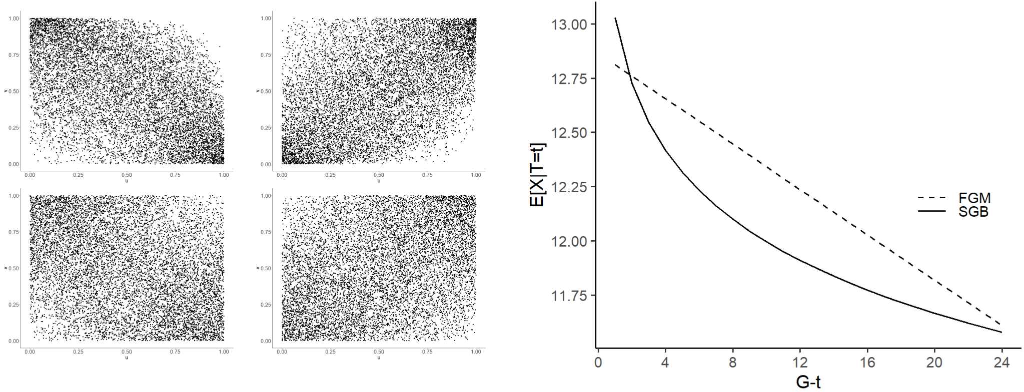


Abbildung: Left: Scatterplots for copulas GB ($\vartheta = 1$), SGB ($\vartheta = 1$), FGM ($\vartheta = 1$ and $\vartheta = -1$); Right: Expected lifespan conditional on age-at-study-begin for SGB and FGM

Table of contents

4 Kolmogorov-Smirnov type test

- Replace (A2) by independence copula $C_\Pi(u, v) := u \cdot v$.
- H_0 : The distribution of $(X_i, T_i)^\mathsf{T}$ comes from $\{F_\theta^X \cdot F^T \mid \theta \in \Theta\}$.
- Truncated point process $N_{n,D}(\cdot) = \sum_{i=1}^n \epsilon_{\binom{X_i}{T_i}}(\cdot \cap D)$
- Estimates:
  - Parameter $\hat{\theta}_n$ as Z-estimator

  $\hookrightarrow$ consistency and asymptotic normality from W & Wied (2022)
  - Sample size $\hat{n} = M_n/\alpha_{\hat{\theta}_n}$ with $M_n = N_{n,D}(D)$

  $\hookrightarrow$ LLN: $M_n/n \xrightarrow{P} \alpha_{\theta_0}$, CLT: $\sqrt{n}\left(\frac{1}{n}M_n - \alpha_{\theta_0}\right) \xrightarrow{d} \mathcal{N}\left(0, \alpha_{\theta_0}(1 - \alpha_{\theta_0})\right)$

4.1 Asymptotic distribution

- Test statistic:

$$\sqrt{\hat{n}} \sup_{(x,t)\in S} \left| \frac{1}{\hat{n}} N_{n,D}([0,x]\times[0,t]) - \mathbb{P}_{\hat{\theta}_n}\left((X_1, T_1)^{\mathsf{T}} \in [0,x]\times[0,t]\cap D\right) \right|$$

- Idea: decomposition into

$$-\underbrace{\sqrt{n}\left(\frac{1}{n}N_{n,D} - \mathbb{P}_{\theta_0}\right)}_{(1)} + \underbrace{\sqrt{n}\left(\frac{1}{n}N_{n,D} - \mathbb{P}_{\hat{\theta}_n}\right)}_{(2)} + \underbrace{\sqrt{\hat{n}}\left(\frac{1}{\hat{n}}N_{n,D} - \mathbb{P}_{\theta_0}\right)}_{(3)}$$

$$+\underbrace{\sqrt{n}\left(\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{\alpha_{\theta_0}}\right)}_{(4)} \underbrace{\frac{1}{\sqrt{\alpha_{\hat{\theta}_n}}}\left[\frac{1}{\sqrt{\hat{n}}}\frac{1}{\sqrt{n}}N_{n,D} + \frac{\sqrt{\alpha_{\theta_0}}}{\sqrt{\alpha_{\hat{\theta}_n}}}\mathbb{P}_{\hat{\theta}_n}\right]}_{(5)} + \text{remainder}$$

- Overview of the four scenarios and limiting processes ($n \to \infty$)

	Test Statistic	Gaussian Process	
		Symbol	Representation
(1)	$\sqrt{n}d(\theta_0, n)$	$\mathbb{B}_{\mathbb{P}_{\theta_0}}$	$\mathbb{B}_{\mathbb{P}_{\theta_0}}$
(2)	$\sqrt{n}d(\hat{\theta}_n, n)$	$\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}$	$\mathbb{B}_{\mathbb{P}_{\theta_0}} - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \dot{\mathbb{E}}_{\theta_0}$
(3)	$\sqrt{\hat{n}}d(\theta_0, \hat{n})$	$\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}$	$\mathbb{B}_{\mathbb{P}_{\theta_0}} - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D) \cdot \mathbb{E}_{\theta_0}/\alpha_{\theta_0}$
(1)-(5)	$\sqrt{\hat{n}}d(\hat{\theta}_n, \hat{n})$	$\mathbb{H}_{\mathbb{P}_{\theta_0}}$	$\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}} + \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \left[\frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}}\mathbb{E}_{\theta_0} - \dot{\mathbb{E}}_{\theta_0}\right]$

- Asympt. linearity $\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\theta_0}(X_i, T_i) + o_{\mathbb{P}_{\theta_0}}(1)$
- Convergence in sup-norm due to continuous mapping theorem

4.2 Calculating the test statistic

- Method according by Justel, Peña, Zamar (1997)
- Evaluation test statistic on a finite set in spite of infinite jumps

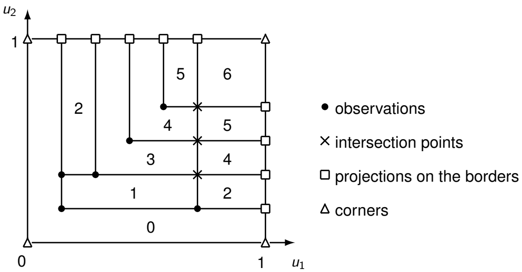


Abbildung: Two dimensional empirical distribution function (multiplied with 6)

4.3 Homogeneity of enterprise lifespans

Test statistic:

- 55 thousend obs.; 22 thousend margin projections; 5.97 mio. intersection points
- Independence copula: 1.65 (largest deviation in intersections)

Quantile:

- Estimator for independent trunkation $\hat{\theta}_n = 0{,}0826$ by W & Wied (2022)
- Grid width $0.01 \Rightarrow$ Number of entries in Covariance mnatrix $(100G \cdot 100(G+s))^2 = 4.2 \cdot 10^{13}$

- Independence copula: 1.65
- Comparison with quantiles:

	C_Π			
Stages	0.10	0.05	0.01	0.10
(θ_0, n)	0.5745	0.6446	0.7805	0.579
$(\hat{\theta}_n, n)$	0.5719	0.6558	0.8298	0.732
$(\theta_0, \hat{n})$	0.3820	0.4235	0.5034	0.386
$(\hat{\theta}_n, \hat{n})$	0.2869	0.3051	0.3558	0.237

$\Rightarrow$ Reject hypothesis

- Independence copula: 1.65 – FGM-Copula: 1.69
- Comparison with quantiles:

	C_Π			C_ϑ^{FGM}		
Schätzgrößen	0.10	0.05	0.01	0.10	0.05	0.01
(θ_0, n)	0.5745	0.6446	0.7805	0.5799	0.6518	0.7862
$(\hat{\theta}_n, n)$	0.5719	0.6558	0.8298	0.7320	0.8510	1.0816
$(\theta_0, \hat{n})$	0.3820	0.4235	0.5034	0.3868	0.4275	0.5092
$(\hat{\theta}_n, \hat{n})$	0.2869	0.3051	0.3558	0.2378	0.2535	0.2914

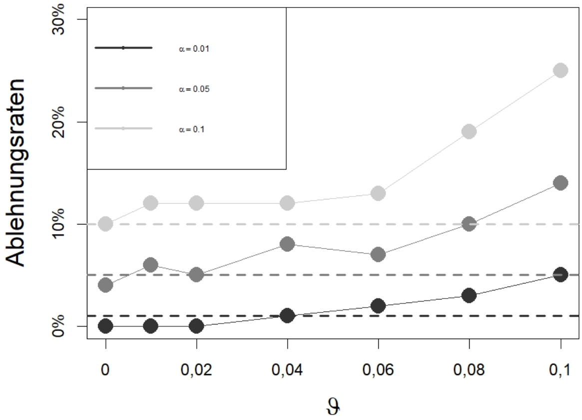
$\Rightarrow$ Reject hypothesis

Table of contents

5 Summary

- Truncation dependence modelling with copulas
  - ⇒ Dependence in business data is weak but significant
  - ↪ multi causal: covariates in micro- or macro-economic sense
  - ↪ +censoring in Sieg/Toparkus/W SPL Nov. 2025
- Limit distribution of truncated process is a Gaussian process
  - – thus construction of Kolmogorov-Smirnov type test
  - ⇒ Oversimplified assumptions on marginals for business data
  - ↪ less frequent business foundations, other lifespan distribution

Simulation of power



Bibliography

- A. Justel, D. Peña, R. Zamar: *A multivariate Kolmogorov-Smirnov test of goodness of fit*. Stat. Probab. Lett. 35, 251-259 (1997)
- P. A. P. Moran: *Maximum-Likelihood estimation in non-standard conditions*. Math. Proc. Camb. Philos. Soc. 70, 441-450 (1971)
- R.-D. Reiss: *A Course an Point Processes*. Springer-Verlag, New York (1993)
- A. Toparkus, R. Weißbach: *Testing truncation dependence: The Gumbel-Barnett Copula*. J. Stat. Plan. Inference 234, 106194 (2024)
- A. W. van der Vaart: *Asymptotic Statistics*. Camb. Univ. Pr., New York (2000)
- R. Weißbach, D. Wied: *Truncating the exponential with a uniform distribution*. Stat Papers 63, 1247-1270 (2022)

Thank you for your patience!