Maastricht University | universiteit hasselt

2015 | Faculty of Sciences

DOCTORAL DISSERTATION

Flexible modeling of clustered event times through frailties and copulas

Doctoral dissertation submitted to obtain the degree of
Doctor of Science: Statistics, to be defended by

Candida Geerdens

Promoter: Prof. Dr Paul Janssen

D/2015/2451/16

universiteit hasselt | I-BioStat | KU LEUVEN

Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Acknowledgements

The realization of this dissertation would not have been possible without the direct or indirect support of many people. Even though I cannot list everyone's name, some people should not go unmentioned.

A sincere word of gratitude is extended to my supervisor Prof. Dr Paul Janssen. Thank you for the excellent guidance and the continuous encouragement. I could not wish for a better promotor! A genuine word of appreciation goes to Prof. Dr Gerda Claeskens and Prof. Dr Noël Veraverbeke. Thank you for the stimulating discussions and the valuable suggestions. I enjoyed working with both of you! I would also like to thank all the committee members for the careful reading and the constructive comments.

Further, this doctoral adventure would have been so much harder if not for the members of the coffee group and the fellow assistants. Thank you all for creating such a pleasant working environment. I value it a lot! Ruth and Leen, it has been a pleasure to share an office with both of you. Thomas, the little walks and the numerous talks will last in memory.

To my parents, thank you for standing beside me during the highs and the lows, for believing in me and for the never failing love.

Candida Geerdens
Diepenbeek, 29 May 2015

Publications

This dissertation is based on the following publications and reports:

Geerdens, C., Claeskens, G. and Janssen, P. (2013). Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics*, 14, 433-446.

Geerdens, C., Claeskens, G. and Janssen, P. (2014). Copula based flexible modeling of associations between clustered event times. Submitted for publication.

Geerdens, C., Janssen, P. and Veraverbeke, N. (2015). Large sample properties of nonparametric copula estimators under bivariate censoring. Submitted for publication.

Contents

List of Tables

List of Figures

List of abbreviations

The abbreviations used in this dissertation are listed below.

AIC	:	Akaike Information Criterion
BIC	:	Bayesian Information Criterion
EAC	:	Exchangeable Archimedean Copula
FNAC	:	Fully Nested Archimedean Copula
IC	:	Information Criterion
i.i.d.	:	independent and identically distributed
NAC	:	Nested Archimedean Copula
PNAC	:	Partially Nested Archimedean Copula
RMSE	:	Root Mean Squared Error

Chapter 1

Introduction

In a survival study the variable of interest is the time to a pre-specified event (e.g., the time to tumor appearance). Often, the event time is right-censored for some study items, i.e., only a lower time bound for the event is observed (e.g., due to drop-out). Further, a study can involve grouped data (e.g., twins - clusters of size two - are followed up for tumor appearance). Since grouped study items chair common traits, their event times exhibit within-cluster correlation.

Popular survival models that account for the association in grouped time-to-event data are the frailty and the copula model. A frailty model is a hazards model supplemented with a cluster-specific random term, named the frailty. Hence, a frailty model is a conditional model. A copula model describes the joint survival function of the event times via the marginal survival functions and a dependence function, called the copula. Typically, a copula model is used for clusters with small and equal size, whereas a frailty model can also handle clusters with large and/or varying size.

In Section 1.1 the basic quantities of survival analysis are reviewed. In Section 1.2 the data sets used to illustrate some of the developed methods are discussed. An overview of the thesis objectives is given in Section 1.3.

1.1 Review of survival analysis concepts

In this section some basic notions on univariate right-censored event time data are introduced.

Let n denote the number of study items. For item r ($r = 1, \ldots, n$) we observe $Y_r = \min(T_r, C_r)$ where T_r is the event time and C_r is the censoring time. The indicator $\delta_r = I(T_r \leq C_r)$ equals one for an event and zero otherwise. Event times and censoring times are assumed to be independent.

Let f be the probability density function and $F(t) = P(T \leq t)$ the cumulative distribution function of T. Basic quantities used to describe time-to-event data are the survival function

$$S(t) = P(T > t) = 1 - F(t)$$

and the hazard function

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

The survival function $S(t)$ gives the probability that a study item survives beyond time t, while the hazard function $h(t)$ represents the instantaneous failure rate given that the item has survived up to time t. A related quantity is the cumulative hazard function

$$H(t) = \int_0^t h(v)dv = -\log S(t).$$

The classical nonparametric estimator for the survival function S is the one introduced by Kaplan and Meier (1958). For all values t within the range of the observed event times, the Kaplan-Meier estimator is defined as

$$\widehat{S}(t) = \prod_{r:Y_{(r)} \leq t} \Big(\frac{n-r}{n-r+1}\Big)^{\delta_{(r)}}$$

where $Y_{(1)} \leq \ldots \leq Y_{(n)}$ are the order statistics of $Y_1, \ldots, Y_n$ and $\delta_{(1)}, \ldots, \delta_{(n)}$ are the corresponding indicators. The Kaplan-Meier curve is a step function with jumps at the observed event times. The stochastic size of the jump at an event time depends on the number of events at that time as well as on the pattern of the censored observations prior to that time. In absence of censoring the Kaplan-Meier estimator reduces to the empirical survival function.

1.2 Data description

In this section we introduce the data sets used to illustrate the methodology developed in the future chapters.

1.2.1 Udder infection data

Mastitis is a bacterial infection of the udder of a dairy cow, which affects the milk production and the milk quality. Depending on the agent infecting the udder, mastitis can even be lethal. Since the infection times in the four udder quarters of a cow are likely to be correlated, we investigate the time to infection taking into account the existing association. A cow is the cluster, the infection times of the four udder quarters are the grouped data.

We consider two sets of mastitis data. In the first data set, 100 cows are followed up for infection. A distinction is made between cows that have had one calving, i.e., one lactation period (primiparous cows - parity $= 1$) and cows that have had multiple calvings (multiparous cows - parity $= 0$). In the second data set, 407 primiparous cows are followed up for infection. The two sets of data come from different studies, the small data set is thus no subset of the larger data set.

An observation is right-censored if no infection occurs before the end of a lactation period, which is roughly $300 - 350$ days but different for every cow, or if a cow is lost to follow-up during the study, e.g., due to culling. Censoring occurs at the level of the cluster (a cow), i.e., the same censoring time applies to the four udder quarters of a cow. However, it may happen that one or more udder quarters are infected (event) with the remaining ones free of infection (censored). In the first data set censoring varies from 19% to 22% in the four udder quarters, overall 21% of the observations are censored. In the second data set censoring ranges from 64% to 69% in the four udder quarters, overall 66% of the observations are censored.

Table 1.1 contains a few lines from the first data set (version 1), while Table 1.2 gives some lines from the second data set (version 2). It follows that, e.g., the infection times of cow 1 in data set 1 are all censored, whereas the infection times of cow 1 in data set 2 are all observed. Similar data have been analyzed by Massonnet *et al.* (2009) and Goethals *et al.* (2009).

1.2.2 Insemination data

In a dairy farm one often registers the time from parturition to first insemination as this is an important factor influencing the length of the calving interval, i.e., the time between two calvings, which optimally lies between 12 and 13 months. Since the insemination policy is similar within a farm, we model the time to first insemination taking into

Table 1.1: Udder infection data - version 1. The first column contains the cow identification number. The second column gives the minimum of the infection time and the censoring time (in days) in each of the four udder quarters. The third column lists the corresponding censoring indicators: one if infected and zero otherwise. The last column reveals the parity of the cow: one for a primiparous cow and zero for a multiparous cow.

cow-id	time	status	parity
1	(308.5, 308.5, 308.5, 308.5)	(0, 0, 0, 0)	1
2	(95, 124, 231, 273)	(1, 1, 1, 1)	1
...	...	...	...
99	(49.5, 49.5, 49.5, 49.5)	(1, 1, 1, 1)	0
100	(159, 233, 83.5, 201.5)	(1, 0, 1, 1)	1

Table 1.2: Udder infection data - version 2. The first column contains the cow identification number. The second column gives the minimum of the infection time and the censoring time (in days) in each of the four udder quarters. The third column reveals the censoring status: one if infected and zero otherwise. Only primiparous cows are considered.

cow-id	time	status
1	(67, 67, 119, 67)	(1, 1, 1, 1)
2	(124, 333, 333, 333)	(1, 0, 0, 0)
...	...	...
406	(220, 220, 220, 220)	(0, 0, 0, 0)
407	(279, 279, 279, 263)	(0, 0, 0, 1)

account the association induced by residence in the same farm. A farm is the cluster, the time to first insemination of the cows in the farm are the grouped data. The data set contains 181 farms, the number of cows per farm varies from 1 to 174 with on average 58 cows.

An observation is right-censored if a cow is not inseminated 300 days after calving or if a cow is lost to follow-up during the study, e.g., due to culling. Censoring does not occur at the level of the cluster (a farm) but at the level of the items in a cluster, i.e., distinct censoring times apply to cows within the same farm. Censoring in a farm ranges from 0% to 100%, overall 18% of the observations are censored.

Table 1.3: Insemination data. Column one and two contain the identification number of the farm, resp. the cow. The third column gives the minimum of the time to first insemination and the censoring time (in days). The fourth column contains the corresponding censoring indicator: one if inseminated and zero otherwise. The last column shows the parity: one for a primiparous cow and zero for a multiparous cow.

farm-id	cow-id	time	status	parity
1	1	68.5	1	0
...	...	...	...	...
1	51	70.5	1	1
...	...	...	...	...
181	10433	48.5	1	0
...	...	...	...	...
181	10513	155.5	1	1

As in the udder infection data parity is included as the single binary covariate. A subset of the data is contained in Table 1.3. It follows that, e.g., cow 1 of farm 1 is multiparous and inseminated at time 68.5, whereas cow 51 of farm 1 is primiparous and inseminated at time 70.5. Similar data have been analyzed by Duchateau and Janssen (2004) and Duchateau *et al.* (2005).

1.3 Outline

The main objective of this dissertation is to develop methods that allow flexible modeling of the association present in clustered right-censored event time data.

Part I concentrates on the frailty model. Chapter 2 contains a brief review on the proportional hazards model and the shared frailty model. In Chapter 3 we construct a test to verify the aptness of an one-parameter gamma frailty density. Based on an orthonormal polynomial expansion, a new class of extended gamma frailty densities is defined. We obtain an explicit expression of the corresponding marginal likelihood for right-censored event time data. Next, we apply an order selection test to find the best fitting model within a considered series of expanded densities. A bootstrap is used to obtain an approximate p-value. The developed method is investigated in a simulation study and applied to the udder infection data as well as the insemination data. The results in Chapter 3 are published in Geerdens *et al.* (2013).

In Part II the focus is on copula modeling. Chapter 4 contains some basic concepts of copula theory. In Chapter 5 exchangeable and nested Archimedean copulas are compared to the more flexible Joe-Hu copula family, which consists of mixtures of max-infinitely divisible bivariate copulas (Joe and Hu, 1996). A likelihood approach is used to fit the diverse copulas. Next, we address the question of model selection. For right-censored time-to-event data, we state conditions under which a penalized likelihood based information criterion is weakly consistent or consistent. The developed method is used in a simulation study and applied to the udder infection data. The material of Chapter 5 is in Geerdens *et al.* (2014). In Chapter 6 a new nonparametric copula estimator is defined for event time data subject to univariate or copula right-censoring. We prove consistency and establish an asymptotic i.i.d. representation. A simulation study is used to investigate the finite sample performance of the proposed estimator as compared to the recent nonparametric copula estimator by Gribkova and Lopez (2014). The results of Chapter 6 are in Geerdens *et al.* (2015).

Part I

The frailty model

Chapter 2

The proportional hazards and the shared frailty model

The proportional hazards model is widely used to describe univariate time-to-event data (Therneau and Grambsch, 2000; Klein and Moeschberger, 2003). For grouped event time data, the within-cluster association needs to be taken into account and hence a modification is required. A cluster-specific random effect is introduced, resulting in the shared frailty model (Duchateau and Janssen, 2008; Wienke, 2011).

In Section 2.1 proportional hazards models are reviewed, shared frailty models are defined in Section 2.2.

2.1 The proportional hazards model

Univariate time-to-event data are often described by means of the proportional hazards model. Here, the hazard function at time t for study item r $(r = 1, \ldots, n)$ with covariate vector $x_r = (x_{r1}, \ldots, x_{rp})'$ of length p is expressed as

$$h_r(t) = h_0(t) \exp(\beta' x_r) \tag{2.1}$$

with $h_0(t)$ the baseline hazard function at time t, i.e., the hazard function at time t of an item whose covariate values equal 0 and β the regression coefficients associated with x_r.

With $h_r(t)$ and $h_l(t)$ the hazard function of the rth, resp. the lth study item at time t, the ratio of two hazard functions is given by

$$\frac{h_r(t)}{h_l(t)} = \exp\left(\beta^{'}(x_r - x_l)\right).$$

The above implies that the hazard ratio of two study items is constant over time, explaining the notion "proportional hazards model".

In model (2.1), h_0 can take a parametric form or can be left unspecified. A popular choice is $h_0(t) = \lambda\rho t^{\rho-1}$ with $\lambda > 0$ a scale parameter and $\rho > 0$ a shape parameter, the Weibull baseline hazard function.

2.2 The shared frailty model

For clustered event time data the proportional hazards model in (2.1) is extended to the shared frailty model, which is given by

$$\begin{aligned} h_{sr}(t) &= h_0(t)\,\exp(\beta^{'}x_{sr} + w_s) \\ &= h_0(t)\,u_s\,\exp(\beta^{'}x_{sr}) \end{aligned} \tag{2.2}$$

where $h_{sr}(t)$ is the hazard function at time t for the rth item $(r = 1,\ldots,n_s)$ in cluster s $(s = 1,\ldots,n)$ given w_s, resp. $u_s = \exp(w_s)$, $h_0(t)$ is the baseline hazard function at time t and β are the regression coefficients associated with the covariate vector $x_{sr} = (x_{sr1},\ldots,x_{srp})^{'}$ of length p. Due to the inclusion of w_s or u_s, a shared frailty model is a conditional model. As in a proportional hazards model, h_0 can take a parametric form (e.g., Weibull) or can be left unspecified.

The factor $u_s = \exp(w_s)$ is the frailty for cluster s $(s = 1,\ldots,n)$ and it is common to all members of that cluster, explaining the concept "shared frailty". It accounts for the similarity in event time of items within the same cluster (correlation). The frailties $u_1,\ldots,u_n$ are assumed to be an i.i.d. sample from a density f_U, where Var(U) represents the heterogeneity between clusters. Common choices for the frailty density are:

- The one-parameter gamma density:

$$f_U(u) = \frac{u^{1/\theta-1}\exp(-u/\theta)}{\theta^{1/\theta}\Gamma(1/\theta)} \tag{2.3}$$

  with $\theta > 0$. The mean and the variance of the frailty are: E$(U) = 1$ and Var$(U) = \theta$.

- The inverse Gaussian density:

$$f_U(u) = \left(\frac{\alpha}{2\pi}\right)^{1/2} u^{-3/2} \exp\left(-\frac{\alpha}{2u\mu^2}(u-\mu)^2\right)$$

  with $\mu > 0$ and $\alpha > 0$. The mean and the variance of the frailty are: $\mathsf{E}(U) = \mu$ and $\mathsf{Var}(U) = \mu^3/\alpha$.

- The positive stable density:

$$f_U(u) = -\frac{1}{\pi u}\sum_{k=1}^{\infty}\frac{\Gamma(k\nu+1)}{k!}(-u^{-\nu})^k \sin(\nu k\pi)$$

  with $0 < \nu < 1$. This density has infinite mean and therefore the variance is also undetermined.

Note that a frailty has a multiplicative effect on the baseline hazard function. To avoid identifiability problems, the mean of the frailty needs to be specified. A standard choice, for frailties with finite mean, is to set the mean equal to one. We illustrate the identifiability issue with an example.

Example. Consider a frailty $\widetilde{U}$ from a density $f_{\widetilde{U}}$ with $\mathsf{E}(\widetilde{U}) \neq 1$ and a frailty $U = \widetilde{U}/\mathsf{E}(\widetilde{U})$ from a density f_U with $\mathsf{E}(U) = 1$. Suppose a Weibull baseline hazard function is appropriate. We have

$$\begin{aligned} h_{sr}(t) &= \lambda\rho t^{\rho-1}\,\widetilde{u}_s\,\exp(\beta' x_{sr}) \\ &= \lambda\rho t^{\rho-1}\,u_s\,\mathsf{E}(\widetilde{U})\,\exp(\beta' x_{sr}) \\ &= \widetilde{\lambda}\rho t^{\rho-1}\,u_s\exp(\beta' x_{sr}) \end{aligned}$$

where $\widetilde{\lambda} = \lambda\mathsf{E}(\widetilde{U})$. We also have $h_{sr}(t) = \lambda\rho t^{\rho-1}\,u_s\,\exp(\beta' x_{sr})$. However, since $\mathsf{E}(\widetilde{U}) \neq 1$, we obtain $\widetilde{\lambda} \neq \lambda$. The Weibull baseline hazard is thus identified differently for both frailty densities.

The strength of the correlation is usually expressed by Kendall's tau (Duchateau and Janssen, 2008). For the one-parameter gamma density $\tau = \theta/(\theta+2)$, while $\tau = 1-\nu$ for the positive stable density. It holds that $0 < \tau < 1$. For the inverse Gaussian density with $\mu = 1$, Kendall's tau equals $0.5 - \alpha + 2\alpha^2\exp(2\alpha)\int_{2\alpha}^{\infty} u^{-1}\exp(-u)du$ and $0 < \tau < 0.5$.

Due to its mathematical convenience, the one-parameter gamma is the most popular density choice. A strategy to verify its aptness is the subject of Chapter 3.

Chapter 3

A goodness-of-fit test for the shared frailty model

As mentioned in Chapter 2, the shared frailty model can be used to describe clustered time-to-event data. The shared frailty model has three components, namely the baseline hazard function, the frailty and the exponential function describing the linear impact of one or more covariates. Due to its mathematical tractability, the frailty is often presumed to follow an one-parameter gamma density. Since it is the frailty density that dictates the type of association, it is essential to verify the adequacy of the chosen density. Here, the focus is on a diagnostic measure for the gamma frailty density, assuming that the two other components of the shared frailty model are correctly specified.

To check the aptness of a gamma frailty density, Shih and Louis (1995) propose, for frailty models with a parametric baseline hazard function, a graphical test based on the evolution over time of the conditional frailty expectation given the observable data. They show that, if a gamma frailty density is adequate, the considered quantity is constant over time. An extension of the Shih and Louis test to an unspecified baseline hazard function is given in Glidden (1999) and in Cui and Sun (2004).

In this chapter we develop, for shared frailty models with a parametric baseline hazard function, a new test to validate the gamma frailty density. To this end, we rely on ideas in Zhang and Davidian (2001) as well as in Claeskens and Hart (2009). They replace the normal density of the random term in a mixed model by a Hermite series expansion and subsequently verify the claim of normality via an order selection test.

For likelihood based models, the asymptotic distribution of an order selection test is established by Aerts *et al.* (1999).

In Section 3.1 we build, based on an orthonormal series expansion, a new class of extended gamma frailty densities. Section 3.2 contains the corresponding marginal likelihood for right-censored event time data. In Section 3.3 we use an order selection test to find the best fitting density within a series of expanded models. In Section 3.4 a simulation study is used to investigate the finite sample performance of the method. In Section 3.5 we analyze the udder infection data and the insemination data. Conclusions and a further discussion are in Section 3.6.

3.1 A class of extended gamma frailty densities

To test the gamma density assumption in a shared frailty model, we define a new class of frailty densities based on polynomials orthonormal to the one-parameter gamma density. Using the Gram-Schmidt orthogonalization procedure we obtain explicit expressions for these polynomials, see Lemma 3.1.1.

Lemma 3.1.1. *For $n = 0, 1, 2, \ldots$ the polynomial functions $p_n : \mathbb{R}^+ \to \mathbb{R} : u \to p_n(u)$ with*

$$\begin{aligned} p_n(u) &= (-\theta)^n \exp(u/\theta) u^{-1/\theta+1} \frac{d^n}{du^n} \left(\exp(-u/\theta) u^{1/\theta+n-1} \right) \\ &= \sum_{i=0}^{n} (-\theta)^{n-i} \binom{n}{i} \frac{\Gamma(1/\theta+n)}{\Gamma(1/\theta+i)} u^i \end{aligned}$$

are orthogonal with respect to the inner product $\langle g_1, g_2 \rangle = \int_0^\infty g_1(u) g_2(u) f_U(u) du$ where f_U is the one-parameter gamma density in (2.3)*. The set of polynomial functions $v_n : \mathbb{R}^+ \to \mathbb{R} : u \to v_n(u)$ with $v_n(u) = p_n(u)/\|p_n\|^{1/2}$ and $\|p_n\| = \theta^{2n} n! \frac{\Gamma(1/\theta+n)}{\Gamma(1/\theta)}$ is orthonormal.*

Proof. Let k be a nonnegative integer. Integrate by parts k times in:

$$\begin{aligned} \int_0^\infty f_U(u) u^k p_n(u) du &= \frac{(-\theta)^n}{\theta^{1/\theta} \Gamma(1/\theta)} \int_0^\infty u^k \frac{d^n}{du^n} \left(\exp(-u/\theta) u^{1/\theta+n-1} \right) du \\ &= \frac{(-\theta)^n}{\theta^{1/\theta} \Gamma(1/\theta)} \Bigg[\underbrace{\left\{ u^k \frac{d^{n-1}}{du^{n-1}} \left(\exp(-u/\theta) u^{1/\theta+n-1} \right) \right\}_0^\infty}_{=0} \\ &\qquad -k \int_0^\infty u^{k-1} \frac{d^{n-1}}{du^{n-1}} \left(\exp(-u/\theta) u^{1/\theta+n-1} \right) du \Bigg] \end{aligned}$$

$$
\begin{aligned}
&= \frac{(-1)^{n+1}\theta^n k}{\theta^{1/\theta}\Gamma(1/\theta)} \int_0^\infty u^{k-1} \frac{d^{n-1}}{du^{n-1}} \left(\exp(-u/\theta)u^{1/\theta+n-1}\right) du \\
&= \dots \\
&= \frac{(-1)^{n+k}\theta^n k!}{\theta^{1/\theta}\Gamma(1/\theta)} \int_0^\infty \frac{d^{n-k}}{du^{n-k}} \left(\exp(-u/\theta)u^{1/\theta+n-1}\right) du.
\end{aligned}
$$

For $k < n$ the above simplifies to

$$
\frac{(-1)^{n+k}\theta^n k!}{\theta^{1/\theta}\Gamma(1/\theta)} \left[\frac{d^{n-k-1}}{du^{n-k-1}} \left(\exp(-u/\theta)u^{1/\theta+n-1}\right)\right]_0^\infty = 0.
$$

Since $p_k(u)$ is a polynomial of degree k, this implies that for $k < n$ it holds that $\int_0^\infty f_U(u)p_k(u)p_n(u)du = 0$. However, for $k = n$ the above simplifies to

$$
\frac{\theta^n n!}{\theta^{1/\theta}\Gamma(1/\theta)} \int_0^\infty \exp(-u/\theta)u^{1/\theta+n-1}du = \theta^{2n} n! \frac{\Gamma(1/\theta+n)}{\Gamma(1/\theta)}
$$

which completes the proof of orthogonality.

Further, since the leading term of $p_n(u)$ is u^n, it follows that

$$
\|p_n\| = \int_0^\infty f_U(u)p_n^2(u)du = \int_0^\infty f_U(u)u^n p_n(u)du = \theta^{2n} n! \frac{\Gamma(1/\theta+n)}{\Gamma(1/\theta)}.
$$

Therefore $v_n(u) = p_n(u)/\|p_n\|^{1/2}$, $n = 0, 1, 2, \dots$ define a system of orthonormal polynomials. □

The first four orthonormal polynomials are: $v_0(u) = 1$, $v_1(u) = (u-1)\theta^{-1/2}$, $v_2(u) = \{u^2 - u(2+2\theta) + \theta + 1\}/\{\theta(2+2\theta)^{1/2}\}$ and $v_3(u) = \{u^3 - 3(1+2\theta)u^2 + 3(1+3\theta+2\theta^2)u - (1+3\theta+2\theta^2)\}/\{6\theta^3(1+3\theta+2\theta^2)\}^{1/2}$. A graphical representation is given in Figure 3.1.

Based on the polynomials in Lemma 3.1.1 an orthonormal series expansion around the one-parameter gamma density f_U is defined as

$$
f_{\widetilde{U}_m}(u) = \frac{f_U(u)}{c(d_{(m)})} \left\{\sum_{j=0}^m d_j v_j(u)\right\}^2 \tag{3.1}
$$

with $\widetilde{U}_m$ the frailty for the model indexed by the series cut-off value m, $d_0 = 1$ and $c(d_{(m)}) = \sum_{j=0}^m d_j^2$ a normalization constant. Note that $m = 0$ corresponds to the one-parameter gamma density f_U and thus $\widetilde{U}_0 = U$. By fitting models with different series cut-off value m (via maximum likelihood estimation) and by using a model selection method to choose the most appropriate value of m, one can verify the aptness of the one-parameter gamma density.

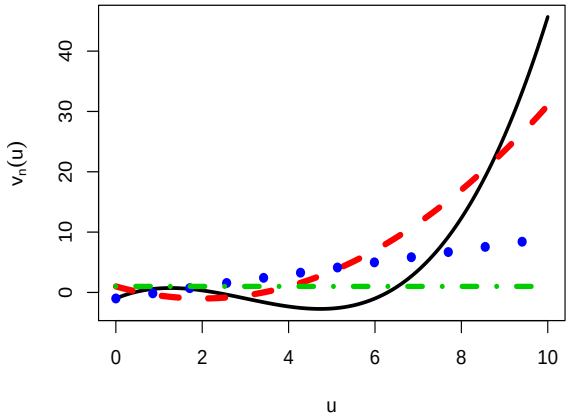


Figure 3.1: Polynomials orthonormal to the one-parameter gamma density with $\theta = 1$; $v_0(u)$ dot-dashed green line, $v_1(u)$ dotted blue line, $v_2(u)$ dashed red line and $v_3(u)$ solid black line.

The class of extended gamma frailty densities defined in (3.1) includes a wide range of alternative frailty densities. In fact, any continuous density on the positive half-line can be approximated by one of its members. Indeed, since $\int_0^\infty \exp(cu) f_U(u)du < \infty$ for any $0 < c < 1/\theta$, the set of orthonormal polynomials v_n $(n = 0, 1, \ldots)$ is closed with respect to continuous functions g on $(0, \infty)$ that satisfy

$$\int_0^\infty g^2(u) f_U(u)du < \infty. \tag{3.2}$$

Consequently the polynomials are complete in the sense that $\lim_{m\to\infty} \int_0^\infty \{g(u) - \sum_{j=0}^m c_j v_j(u)\}^2 f_U(u)du = 0$ for all continuous functions g on $(0, \infty)$ that satisfy (3.2) (Nikiforov and Uvarov, 1988). Take $\widetilde{f}$ any continuous density function on $(0, \infty)$, and define $g(u) = \{\widetilde{f}(u)/f_U(u)\}^{1/2}$. It is readily verified that (3.2) holds and hence the series expansion in (3.1) is able to describe any continuous density on the positive real line, where $c_j = \int_0^\infty \{\widetilde{f}(u) f_U(u)\}^{1/2} v_j(u)du$. If moreover $\int_0^\infty \{g'(u)\}^2 u f_U(u)du < \infty$, then the series expansion $g(u) = \sum_{j=0}^\infty c_j v_j(u)$ converges uniformly on every interval $[u_1, u_2] \subset (0, \infty)$ (Nikiforov and Uvarov, 1988).

Next, the mean value of the frailty $\widetilde{U}_m$ is investigated. To this end, define $a_{ij} = (-\theta)^{j-i}\binom{j}{i}\frac{\Gamma(1/\theta+j)}{\Gamma(1/\theta+i)}$ and $a^*_{ij} = a_{ij}/\|p_j\|^{1/2}$ as well as $b^*_{ij} = a^{*2}_{ij} + 2\sum_{k+l=2i, k<l\leq j} a^*_{kj} a^*_{lj}$ and $c^*_{ij} = 2\sum_{k+l=2i+1, k<l\leq j} a^*_{kj} a^*_{lj}$.

Lemma 3.1.2. *For $\widetilde{U}_m$ a random variable with density $f_{\widetilde{U}_m}$ as in* (3.1),

$$E_m = E(\widetilde{U}_m) = \left[\sum_{j=0}^{m} d_j^2 \left\{ \sum_{i=0}^{j} b_{ij}^* \theta^{2i+1} \Gamma(2i+1/\theta+1) + \sum_{i=0}^{j-1} c_{ij}^* \theta^{2i+2} \Gamma(2i+1/\theta+2) \right\} + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \left\{ \sum_{i=0}^{j} \sum_{l=0}^{k} a_{ij}^* a_{lk}^* \theta^{i+l+1} \Gamma(i+l+1/\theta+1) \right\} \right] \frac{1}{c(d_{(m)})\Gamma(1/\theta)}.$$

Proof. Note that for $j \neq k$, $v_j(u)v_k(u) = \sum_{i=0}^{j} \sum_{l=0}^{k} a_{ij}^* a_{lk}^* u^{i+l}$ and $v_j^2(u) = \sum_{i=0}^{j} b_{ij}^* u^{2i} + \sum_{i=0}^{j-1} c_{ij}^* u^{2i+1}$. The mean frailty can then be calculated as

$$\begin{aligned}
\mathsf{E}(\widetilde{U}_m) &= \int_0^\infty u f_{\widetilde{U}_m}(u)du = \int_0^\infty u \frac{f_U(u)}{c(d_{(m)})} \left[\sum_{j=0}^{m} d_j v_j(u)\right]^2 du \\
&= \int_0^\infty u \frac{f_U(u)}{c(d_{(m)})} \left[\sum_{j=0}^{m} d_j^2 v_j^2(u) + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k v_j(u) v_k(u)\right] du \\
&= \frac{1}{c(d_{(m)})} \left[\sum_{j=0}^{m} d_j^2 \int_0^\infty u f_U(u) v_j^2(u) du + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \int_0^\infty u f_U(u) v_j(u) v_k(u) du\right] \\
&= \frac{1}{c(d_{(m)})} \left[\sum_{j=0}^{m} d_j^2 \int_0^\infty u f_U(u) \left\{ \sum_{i=0}^{j} b_{ij}^* u^{2i} + \sum_{i=0}^{j-1} c_{ij}^* u^{2i+1} \right\} du \right. \\
&\qquad \left. + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \int_0^\infty u f_U(u) \sum_{i=0}^{j} \sum_{l=0}^{k} a_{ij}^* a_{lk}^* u^{i+l} du\right] \\
&= \frac{1}{c(d_{(m)})} \left[\sum_{j=0}^{m} d_j^2 \int_0^\infty f_U(u) \left\{ \sum_{i=0}^{j} b_{ij}^* u^{2i+1} + \sum_{i=0}^{j-1} c_{ij}^* u^{2i+2} \right\} du \right. \\
&\qquad \left. + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \int_0^\infty f_U(u) \sum_{i=0}^{j} \sum_{l=0}^{k} a_{ij}^* a_{lk}^* u^{i+l+1} du\right] \\
&= \frac{1}{c(d_{(m)})\theta^{1/\theta}\Gamma(1/\theta)} \left[\sum_{j=0}^{m} d_j^2 \int_0^\infty \exp(-u/\theta) \left\{ \sum_{i=0}^{j} b_{ij}^* u^{2i+1/\theta} + \sum_{i=0}^{j-1} c_{ij}^* u^{2i+1/\theta+1} \right\} du \right. \\
&\qquad \left. + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \int_0^\infty \exp(-u/\theta) \sum_{i=0}^{j} \sum_{l=0}^{k} a_{ij}^* a_{lk}^* u^{i+l+1/\theta} du\right] \\
&= \frac{1}{c(d_{(m)})\theta^{1/\theta}\Gamma(1/\theta)} \left[\sum_{j=0}^{m} d_j^2 \left\{ \sum_{i=0}^{j} b_{ij}^* \int_0^\infty \exp(-u/\theta) u^{2i+1/\theta} du \right. \right. \\
&\qquad \left. \left. + \sum_{i=0}^{j-1} c_{ij}^* \int_0^\infty \exp(-u/\theta) u^{2i+1/\theta+1} du \right\} \right.
\end{aligned}$$

$$+2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^* \int_0^\infty \exp\left(-u/\theta\right) u^{i+l+1/\theta} du\Bigg].$$

Using the change of variable $t = u/\theta$ as well as the definition of the gamma function, i.e., $\Gamma(a+1) = \int_0^\infty \exp(-u)u^a du$, we have

$$\begin{aligned}
E(\widetilde{U}_m) \;=\; & \frac{1}{c(d_{(m)})\theta^{1/\theta}\Gamma\left(1/\theta\right)}\Bigg[\sum_{j=0}^{m} d_j^2\Bigg\{\sum_{i=0}^{j} b_{ij}^*\theta^{2i+1/\theta+1}\int_0^\infty \exp\left(-t\right) t^{2i+1/\theta} dt \\
& +\sum_{i=0}^{j-1} c_{ij}^*\theta^{2i+1/\theta+2}\int_0^\infty \exp\left(-t\right) t^{2i+1/\theta+1} dt\Bigg\} \\
& +2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^*\theta^{i+l+1/\theta+1}\int_0^\infty \exp\left(-t\right) t^{i+l+1/\theta} dt\Bigg] \\
= \; & \frac{1}{c(d_{(m)})\theta^{1/\theta}\Gamma\left(1/\theta\right)}\Bigg[\sum_{j=0}^{m} d_j^2\Bigg\{\sum_{i=0}^{j} b_{ij}^*\theta^{2i+1/\theta+1}\Gamma\left(2i+1/\theta+1\right) \\
& +\sum_{i=0}^{j-1} c_{ij}^*\theta^{2i+1/\theta+2}\Gamma\left(2i+1/\theta+2\right)\Bigg\} \\
& +2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^*\theta^{i+l+1/\theta+1}\Gamma\left(i+l+1/\theta+1\right)\Bigg] \\
= \; & \frac{1}{c(d_{(m)})\Gamma\left(1/\theta\right)}\Bigg[\sum_{j=0}^{m} d_j^2\Bigg\{\sum_{i=0}^{j} b_{ij}^*\theta^{2i+1}\Gamma\left(2i+1/\theta+1\right) \\
& +\sum_{i=0}^{j-1} c_{ij}^*\theta^{2i+2}\Gamma\left(2i+1/\theta+2\right)\Bigg\} \\
& +2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^*\theta^{i+l+1}\Gamma\left(i+l+1/\theta+1\right)\Bigg].
\end{aligned}$$

□

Note that E_m depends on $d_{(m)}$ as well as on the series cut-off value m. Due to their multiplicative effect on the baseline hazard, we need to standardize the frailties $\widetilde{U}_m$ to have mean equal to one in order for model (2.2) to be well defined. Indeed, otherwise the baseline hazard would change from one model to another (Section 2.2). We thus define the frailty $U_m = \widetilde{U}_m/E_m$. It follows that the density of the standardized frailty equals

$$f_{U_m}(u) = \frac{f_U(uE_m)E_m}{c(d_{(m)})}\left\{\sum_{j=0}^{m} d_j v_j(uE_m)\right\}^2. \tag{3.3}$$

3.2 The marginal loglikelihood

Consider right-censored time-to-event data. Denote by $Y_{sr} = \min(T_{sr}, C_{sr})$ the observed time of item r $(r = 1, \ldots, n_s)$ in cluster s $(s = 1, \ldots, n)$, T_{sr} is the true event time and C_{sr} is the censoring time. The indicator $\delta_{sr} = I(T_{sr} \leq C_{sr})$ equals one if $Y_{sr} = T_{sr}$ and zero otherwise. Event times and censoring times are assumed to be independent.

Let H_0 be the cumulative baseline hazard and x_{sr} the covariate vector of item r in cluster s. Define $A_s = \prod_{r=1}^{n_s} \{h_0(Y_{sr}) \exp(\beta' x_{sr})\}^{\delta_{sr}}$, $B_s = \sum_{r=1}^{n_s} H_0(Y_{sr}) \exp(\beta' x_{sr})$ and $D_s = \sum_{r=1}^{n_s} \delta_{sr}$ $(s = 1, \ldots, n)$. Denote the parameter vector of the baseline hazard by ξ and use ζ_m as notation for the vector $(\xi, \beta, \theta, d_{(m)})$ where $d_{(m)} = (d_0, \ldots, d_m)$.

A nice feature of a shared frailty model with an extended gamma frailty density is that the marginal loglikelihood of the data has a closed form.

Theorem 3.2.1. *The marginal loglikelihood of the extended gamma frailty model with frailty density f_{U_m} as in* (3.3) *is given by*

$$\begin{aligned}\ell_{m,\mathrm{marg}}(\zeta_m) \;=\; & \sum_{s=1}^{n} \log\Bigg(\frac{A_s E_m^{1/\theta}}{c(d_{(m)})\theta^{1/\theta}\Gamma(1/\theta)} \Bigg[\sum_{j=0}^{m} d_j^2 \Bigg\{ \sum_{i=0}^{j} \frac{b_{ij}^* E_m^{2i} \Gamma(2i + D_s + 1/\theta)}{(B_s + E_m/\theta)^{2i+D_s+1/\theta}} \\ & + \sum_{i=0}^{j-1} \frac{c_{ij}^* E_m^{2i+1} \Gamma(2i + D_s + 1/\theta + 1)}{(B_s + E_m/\theta)^{2i+D_s+1/\theta+1}} \Bigg\} \\ & + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j} \sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l} \Gamma(i + l + D_s + 1/\theta)}{(B_s + E_m/\theta)^{i+l+D_s+1/\theta}} \Bigg] \Bigg).\end{aligned}$$

Proof. For cluster s the conditional likelihood, $L_s(\xi, \beta | U_{m,s} = u_s)$, is given by

$$\begin{aligned}& \prod_{r=1}^{n_s} \left[\left\{ h_0(Y_{sr})\, u_s \exp(\beta' x_{sr}) \right\}^{\delta_{sr}} \exp\left(-H_0(Y_{sr})\, u_s \exp(\beta' x_{sr}) \right) \right] \\ & = \prod_{r=1}^{n_s} u_s^{\delta_{sr}} \prod_{r=1}^{n_s} \left\{ h_0(Y_{sr}) \exp(\beta' x_{sr}) \right\}^{\delta_{sr}} \exp\Big(-u_s \sum_{r=1}^{n_s} H_0(Y_{sr}) \exp(\beta' x_{sr}) \Big).\end{aligned}$$

Using the notation A_s, B_s and D_s, the marginal likelihood for the sth cluster, i.e., $L_{m,marg,s}(\zeta_m)$, can be computed as

$$\int_0^\infty L_s(\xi, \beta | U_{m,s} = u_s)\, f_{U_m}(u_s) du_s = A_s \int_0^\infty u_s^{D_s} \exp(-u_s B_s) f_{U_m}(u_s) du_s$$

$$
\begin{aligned}
&= \frac{A_s E_m}{c(d_{(m)})}\left[\int_0^\infty u_s^{D_s}\exp(-u_s B_s) f_U(u_s E_m)\left\{\sum_{j=0}^{m} d_j v_j(u_s E_m)\right\}^2 du_s\right] \\
&= \frac{A_s E_m}{c(d_{(m)})}\left[\int_0^\infty u_s^{D_s}\exp(-u_s B_s) f_U(u_s E_m)\right. \\
&\qquad \left.\left\{\sum_{j=0}^{m} d_j^2 v_j^2(u_s E_m) + 2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k v_j(u_s E_m) v_k(u_s E_m)\right\} du_s\right] \\
&= \frac{A_s E_m}{c(d_{(m)})}\left[\sum_{j=0}^{m} d_j^2 \int_0^\infty u_s^{D_s}\exp(-u_s B_s) f_U(u_s E_m) v_j^2(u_s E_m) du_s\right. \\
&\qquad \left. +2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \int_0^\infty u_s^{D_s}\exp(-u_s B_s) f_U(u_s E_m) v_j(u_s E_m) v_k(u_s E_m) du_s\right] \\
&= \frac{A_s E_m}{c(d_{(m)})}\Big[F_s + G_s\Big].
\end{aligned}
$$

Using the change of variable $t_s = u_s\,(B_s + E_m/\theta)$ as well as the definition of the gamma function $\Gamma(a+1) = \int_0^\infty \exp(-u)u^a du$, the terms F_s and G_s can be rewritten as

$$
\begin{aligned}
F_s &= \sum_{j=0}^{m} d_j^2\left[\int_0^\infty \exp(-u_s B_s) f_U(u_s E_m)\right. \\
&\qquad \left.\left\{\sum_{i=0}^{j} b_{ij}^* E_m^{2i} u_s^{2i+D_s} + \sum_{i=0}^{j-1} c_{ij}^* E_m^{2i+1} u_s^{2i+D_s+1}\right\} du_s\right] \\
&= \frac{E_m^{1/\theta-1}}{\theta^{1/\theta}\Gamma(1/\theta)}\left[\sum_{j=0}^{m} d_j^2\left\{\int_0^\infty \exp(-u_s(B_s + E_m/\theta))\right.\right. \\
&\qquad \left.\left.\sum_{i=0}^{j} b_{ij}^* E_m^{2i} u_s^{2i+D_s+1/\theta-1} + \sum_{i=0}^{j-1} c_{ij}^* E_m^{2i+1} u_s^{2i+D_s+1/\theta}\right\} du_s\right] \\
&= \frac{E_m^{1/\theta-1}}{\theta^{1/\theta}\Gamma(1/\theta)}\left[\sum_{j=0}^{m} d_j^2\left\{\sum_{i=0}^{j}\frac{b_{ij}^* E_m^{2i}}{(B_s + E_m/\theta)^{2i+D_s+1/\theta}}\int_0^\infty \exp(-t_s) t_s^{2i+D_s+1/\theta-1} dt_s\right.\right. \\
&\qquad \left.\left. + \sum_{i=0}^{j-1}\frac{c_{ij}^* E_m^{2i+1}}{(B_s + E_m/\theta)^{2i+D_s+1/\theta+1}}\int_0^\infty \exp(-t_s) t_s^{2i+D_s+1/\theta} dt_s\right\}\right] \\
&= \frac{E_m^{1/\theta-1}}{\theta^{1/\theta}\Gamma(1/\theta)}\left[\sum_{j=0}^{m} d_j^2\left\{\sum_{i=0}^{j}\frac{b_{ij}^* E_m^{2i}\Gamma(2i + D_s + 1/\theta)}{(B_s + E_m/\theta)^{2i+D_s+1/\theta}}\right.\right. \\
&\qquad \left.\left. + \sum_{i=0}^{j-1}\frac{c_{ij}^* E_m^{2i+1}\Gamma(2i + D_s + 1/\theta + 1)}{(B_s + E_m/\theta)^{2i+D_s+1/\theta+1}}\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
G_s &= 2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \Bigg[\int_0^\infty \exp(-u_s B_s) f_U(u_s E_m) \\
&\qquad \left\{ \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^* E_m^{i+l} u_s^{i+l+D_s} du_s \right\} \Bigg] \\
&= \frac{2E_m^{1/\theta-1}}{\theta^{1/\theta}\Gamma\left(1/\theta\right)} \Bigg[\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \Bigg\{ \int_0^\infty \exp(-u_s(B_s + E_m/\theta)) \\
&\qquad \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^* E_m^{i+l} u_s^{i+l+D_s+1/\theta-1} \Bigg\} du_s \Bigg] \\
&= \frac{2E_m^{1/\theta-1}}{\theta^{1/\theta}\Gamma\left(1/\theta\right)} \Bigg[\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l}}{\left(B_s + E_m/\theta\right)^{i+l+D_s+1/\theta}} \\
&\qquad \int_0^\infty \exp(-t_s) t_s^{i+l+D_s+1/\theta-1} dt_s \Bigg] \\
&= \frac{2E_m^{1/\theta-1}}{\theta^{1/\theta}\Gamma\left(1/\theta\right)} \Bigg[\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l} \Gamma\left(i+l+D_s+1/\theta\right)}{\left(B_s + E_m/\theta\right)^{i+l+D_s+1/\theta}} \Bigg].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
L_{m,marg,s}\left(\zeta_m\right) &= \frac{A_s E_m^{1/\theta}}{c(d_{(m)})\theta^{1/\theta}\Gamma(1/\theta)} \Bigg[\sum_{j=0}^{m} d_j^2 \Bigg\{ \sum_{i=0}^{j} \frac{b_{ij}^* E_m^{2i}\Gamma(2i+D_s+1/\theta)}{(B_s+E_m/\theta)^{2i+D_s+1/\theta}} \\
&\qquad + \sum_{i=0}^{j-1} \frac{c_{ij}^* E_m^{2i+1}\Gamma(2i+D_s+1/\theta+1)}{(B_s+E_m/\theta)^{2i+D_s+1/\theta+1}} \Bigg\} \\
&\qquad + 2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l}\Gamma(i+l+D_s+1/\theta)}{(B_s+E_m/\theta)^{i+l+D_s+1/\theta}} \Bigg].
\end{aligned}
$$

Taking the log and summing over the n clusters one gets the expression for the marginal loglikelihood $\ell_{m,marg}\left(\zeta_m\right) = \sum_{s=1}^{n} \log\left(L_{m,marg,s}\left(\zeta_m\right)\right)$ as stated in Theorem 3.2.1. □

3.3 A likelihood ratio order selection test

3.3.1 Null and alternative hypotheses

In model (2.2) we wish to test the null hypothesis that the frailties come from an one-parameter gamma density

$$
\mathcal{H}_0 : U \sim f_U(u) = \frac{u^{1/\theta-1}\exp(-u/\theta)}{\theta^{1/\theta}\Gamma(1/\theta)} \text{ for some value } \theta > 0.
$$

The alternative hypothesis $\mathcal{H}_a$ states that the frailty density is not an one-parameter gamma density. Using models with a different series cut-off value $m = 0, 1, \ldots, M_n$ in the class of extended gamma frailty densities, the hypotheses can be rephrased as

$$\mathcal{H}_0 : m = 0 \text{ which is equivalent to: for all } j = 1, \ldots, M_n : d_j = 0$$
$$\mathcal{H}_a : m > 0 \text{ which is equivalent to: there exists a } j \in \{1, \ldots, M_n\} : d_j \neq 0.$$

Different values of $m = 1, 2, \ldots, M_n$ lead to different extensions of the gamma density. The models based on a large value of m contain the models with a smaller value of m as special cases. In other words, a nested model sequence is constructed by letting m grow. In practice M_n is chosen fixed and not too large, since deviations from $\mathcal{H}_0$ will typically be detected for fixed small values of M_n.

3.3.2 Order selection test

For likelihood based models, Aerts *et al.* (1999) defined the order selection (OS) statistic, rephrased to our setting, by

$$T_{n,OS} = \max_{1 \leq m \leq M_n} 2 \, \frac{\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - \ell_{0,\mathrm{marg}}(\widehat{\zeta}_0)}{m}$$

where ζ_m is the parameter vector of the model using density f_{U_m}. Denote the length of ζ_m by q_m.

With $\widehat{\zeta}_0$ and $\widehat{\zeta}_m$ the maximum likelihood estimators of ζ_0, resp. ζ_m, the numerator of $T_{n,OS}$ coincides with a likelihood ratio test. Further, note that an increase of the series cutoff value m in (3.3) with one, implies the addition of one polynomial v_{m+1} to the series or equivalently the addition of d_{m+1} to the vector of coefficients. Therefore, the denominator of $T_{n,OS}$ equals the difference in number of parameters of model m as compared to the null model, i.e., $q_m - q_0 = m$. The omnibus nature of the test becomes clear. The test is not a single likelihood ratio test, but a maximum of weighted likelihood ratio statistics. By taking a maximum, the statistic $T_{n,OS}$ combines various likelihood ratio statistics and thereby handles the multiple testing issue. The weights $1/m$ take the complexity of the models into account with a down-weight for large models.

Define a modified version of AIC (Akaike, 1973) by $\mathrm{AIC}_{C_\alpha}(m) = 2\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - C_\alpha q_m$, where α denotes a pre-specified significance level. It then follows that rejecting $\mathcal{H}_0$ when $T_{n,OS} > C_\alpha$ corresponds to rejecting $\mathcal{H}_0$ when $m^* = \arg\max_{m=0,\ldots,M_n} \mathrm{AIC}_{C_\alpha}(m) > 0$,

explaining the name "order selection test". Indeed,

$$
\begin{aligned}
& T_{n,OS} > C_\alpha \\
\Leftrightarrow \quad & \max_{1\leq m\leq M_n} 2\,\frac{\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - \ell_{0,\mathrm{marg}}(\widehat{\zeta}_0)}{m} > C_\alpha \\
\Leftrightarrow \quad & \exists m \in \{1,\ldots,M_n\} : 2\,\frac{\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - \ell_{0,\mathrm{marg}}(\widehat{\zeta}_0)}{m} > C_\alpha \\
\Leftrightarrow \quad & \exists m \in \{1,\ldots,M_n\} : 2\,\{\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - \ell_{0,\mathrm{marg}}(\widehat{\zeta}_0)\} > C_\alpha m \\
\Leftrightarrow \quad & \exists m \in \{1,\ldots,M_n\} : 2\,\{\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - \ell_{0,\mathrm{marg}}(\widehat{\zeta}_0)\} > C_\alpha \{q_m - q_0\} \\
\Leftrightarrow \quad & \exists m \in \{1,\ldots,M_n\} : 2\,\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - C_\alpha q_m > 2\,\ell_{0,\mathrm{marg}}(\widehat{\zeta}_0) - C_\alpha q_0 \\
\Leftrightarrow \quad & \exists m \in \{1,\ldots,M_n\} : \mathsf{AIC}_{C_\alpha}(m) > \mathsf{AIC}_{C_\alpha}(0) \\
\Leftrightarrow \quad & m^* = \arg\max_{m=0,\ldots,M_n} \mathsf{AIC}_{C_\alpha}(m) > 0.
\end{aligned}
$$

To obtain the critical value C_α, information on the asymptotic distribution of $T_{n,OS}$ is needed. For a test about the density of the random effect in a mixed model, Claeskens and Hart (2009) show that $T_{OS} = \max_{m\geq 1} \frac{V_m}{m}$, where $V_m = \sum_{j=1}^{m} Z_j^2$ with $Z_1, Z_2, \ldots$ independent $N(0,1)$ distributed random variables, is the limiting form of $T_{n,OS}$ as $n \to \infty$. However, for the hypotheses in Section 3.3.1, a small simulation study based on 1000 data sets, each with 150 clusters of size 4, $\theta = 0.3$ or 0.5 and $M_n = 3$, reveals that the finite sample null distribution of $T_{n,OS}$ depends on the value of θ (see Figure 3.2). Moreover, a concise simulation study based on 1000 data sets, each with 1000 or 5000 clusters of size 2, $\theta = 0.3$ and $M_n = 3$, shows that convergence to the null distribution postulated by Claeskens and Hart (2009) is rather slow. Indeed, according to the latter $C_{0.05} = 4.18$, while based on the simulated distribution, we obtain $C_{0.05} = 2.46$ resp. $C_{0.05} = 3.14$ for the setting with 1000, resp. 5000 clusters. We therefore apply a bootstrap approach to obtain approximate p-values. A detailed description of the bootstrap algorithm for a shared frailty model can be found in Appendix A.1.

3.4 Simulation study

To evaluate the numerical performance of the proposed methodology, we set up a small simulation study. The settings used are inspired by the study on the performance of the gamma frailty model in Duchateau and Janssen (2008).

3.4.1 Simulation setting

Event times T_{sr} are generated based on model (2.2) with, for a binary covariate, a hazard rate of 1.3, i.e., $\beta = \log(1.3)$ and a Weibull baseline hazard defined by a scale

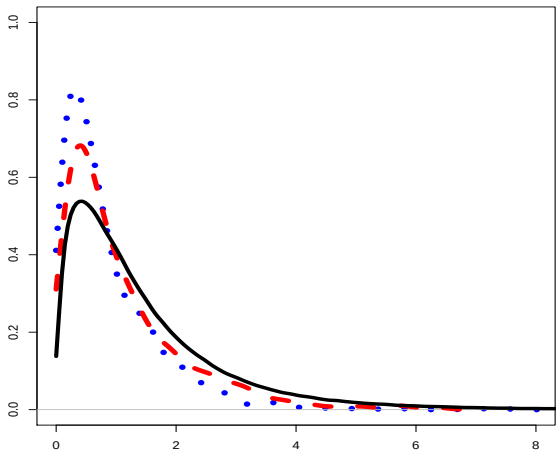


Figure 3.2: Kernel estimates of the finite sample density of $T_{n,OS}$ for true values of θ equal to 0.3 (dotted blue line) and 0.5 (dashed red line). The solid black line represents the asymptotic density of $T_{n,OS}$.

$\lambda = 0.22$ and a shape $\rho = 1$ (exponential baseline hazard). The frailties come from either an one-parameter gamma density with $\theta = 0.3$ (f_1), an inverse Gaussian density with $\mu = 1$ and $\alpha = 10/3$ (variance of 0.3) (f_2), resp. $\mu = 2$ and $\alpha = 5$ (variance of 1.6) (f_3) or from a positive stable density with $\nu = 0.85$ (f_4) (see Section 2.2). The latter three serve as alternative frailty densities. Note that, in terms of mean and variance, f_2 is close to f_1, f_3 is away from f_1 in a moderate way and f_4 is far away from f_1. Also, the corresponding Kendall's tau values are quite similar for the diverse frailty densities: $\tau_1 = 0.13$, $\tau_2 = 0.11$, $\tau_3 = 0.13$ and $\tau_4 = 0.15$. To obtain censoring times C_{sr}, think about a trial where patients enter the study in a uniform way over an accrual period of five years and with a follow up period of three years. The censoring time for a subject then consists of the time at risk before the end of the accrual period plus the follow up time.

Since the maximization of the likelihood is numerically difficult and time-consuming, especially when the value of m is large, which combined with a bootstrap algorithm costs even more time, the investigation is limited to 200 simulated data sets containing 150 or 300 clusters (n) of size $n_s = 4$, each supplemented with 150 bootstrap samples. M_n is taken to be 3. While for a single data example the accuracy can be taken higher, this was not feasible in the simulation study.

The observed data are then given by $Y_{sr} = \min(T_{sr}, C_{sr})$ and $\delta_{sr} = I(T_{sr} \leq C_{sr})$ ($s = 1, \ldots, 150$ or $300, r = 1, \ldots, 4$). Under the gamma frailty model (f_1), we obtain 33% censored observations, while for the inverse Gaussian and positive stable frailty densities, the percentages of censored observations are approximately 32% (f_2), 15% (f_3) and 1.4% for f_4. For a trial with uniform censoring times it is natural that the percentage of censored observations decreases for frailties that take larger (mean) values. Equal censoring percentages can however be obtained by adapting the used censoring mechanism.

The concrete settings we consider for (n,n_s,frailty density) are thus (150,4,f_1), (300,4,f_1), (150,4,f_2), (300,4,f_2), (150,4,f_3), (300,4,f_3), (150,4,f_4), (300,4,f_4). To obtain p-values we use the parametric bootstrap algorithm described in Appendix A.1.

3.4.2 Simulation results

Table 3.1 lists the percentage of times that the null hypothesis is rejected. The level of the test is approximately attained and the positive stable alternative model (f_4) gives the highest simulated rejection probability. For the inverse Gaussian alternative model f_3 the test is also able to pick up the deviation from the null model, but with less pronounced simulated rejection probability. However, for the inverse Gaussian alternative model f_2, the detection more or less fails. Intuition for the simulated rejection probabilities in Table 3.1 can be obtained from the information contained in Figure 3.3. In the left panel data from a shared positive stable frailty model (f_4) with Weibull baseline and one binary covariate is considered. Given these data we fit a shared gamma frailty model, leading to an estimated parameter $\widehat{\theta}$ of 0.20 (se $= 0.21 \cdot 10^{-2}$). This gamma density is depicted together with the positive stable density used to generate the data. The estimated gamma density is in some sense the 'best possible null' frailty for the given data set. It is clear from the picture that the 'best possible null frailty' is far from the 'real' positive stable frailty, which explains the high values for the simulated rejection probabilities. In the right panel we do the same but now for the inverse Gaussian frailty model based on f_3. For this setting the 'best possible null frailty' ($\widehat{\theta} = 0.29$, se $= 0.27 \cdot 10^{-2}$) is closer to the 'real' inverse Gaussian frailty. Therefore, the test is still able to detect the deviation from the null model but, compared to the positive stable alternative model, with lower simulated rejection probability. As expected, the simulated rejection probabilities of the test increase with increased sample size for data generated under the alternative hypothesis.

Table 3.1: Simulation results. The first column reveals the simulation setting. The second and third column list the percentages of simulated data sets for which the null hypothesis is rejected at level $\alpha = 0.05$, resp. at level $\alpha = 0.10$. The first two rows correspond to the null hypothesis (simulated significance levels).

(n,n_s,frailty density)	$\alpha = 0.05$	$\alpha = 0.10$
(150,4,f_1)	0.09	0.14
(300,4,f_1)	0.03	0.09
(150,4,f_2)	0.10	0.15
(300,4,f_2)	0.08	0.15
(150,4,f_3)	0.14	0.22
(300,4,f_3)	0.20	0.28
(150,4,f_4)	0.91	0.92
(300,4,f_4)	0.99	0.99

3.5 Illustrative example

We apply the proposed methodology to the udder infection data described in Section 1.2.1 and the insemination data of Section 1.2.2.

3.5.1 Udder infection data

Consider the udder infection data - version 1 (100 cows). We assume a Weibull baseline hazard, i.e., $\xi = (\lambda, \rho)$. When taking $M_n = 5$ the value of the order selection test statistic $T_{n,OS}$ is 4.82 and the bootstrapped p-value based on 300 samples equals 0.03. We therefore reject the null hypothesis. The value of M_n is not that important, also for, e.g., $M_n = 3$ the conclusion holds. For the given data set, the extended gamma frailty density with $m = 2$ is preferred: with $C_{0.05}$ estimated to be 4.15, $\text{AIC}_{C_{0.05}}(0)$ for the null model equals -841.31 while $m = 2$ corresponds to an $\text{AIC}_{C_{0.05}}(2)$ of -839.98. The parameter estimates of the chosen density are $\widehat{\lambda} = 8.31 \cdot 10^{-5}$ (se $= 4.14 \cdot 10^{-5}$), $\widehat{\rho} = 2.05$ (se $= 0.11$), $\widehat{\beta} = 0.34$ (se $= 0.29$), $\widehat{\theta} = 0.92$ (se $= 0.25$), $\widehat{d_1} = 0.09$ (se $= 0.09$) and $\widehat{d_2} = 0.48$ (se $= 0.16$). Figure 3.4 shows the estimated null density together with the estimated preferred frailty density.

For the udder infection data we plot, in the left panel of Figure 3.5, the centered average of the posterior mean frailty over time together with the 95% confidence limits (Shih and Louis, 1995). The observed pattern is only constant quite late in time. Hence,

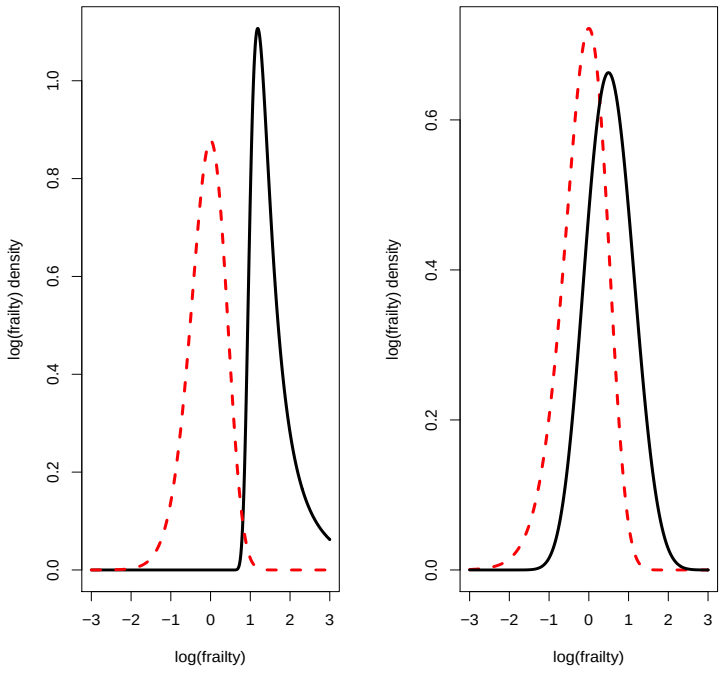


Figure 3.3: Comparison of the true log-frailty density (solid black line) with the estimated log-frailty one-parameter gamma density (dashed red line). Left panel: true density is positive stable with $\nu = 0.85$ and estimated gamma density parameter is $\widehat{\theta} = 0.20$, right panel: true density is inverse Gaussian with $\mu = 2, \alpha = 5$ and estimated gamma density parameter is $\widehat{\theta} = 0.29$.

the gamma frailty density assumption is questionable. This is in line with our finding.

3.5.2 Insemination data

Consider the insemination data. We assume a Weibull baseline hazard, i.e., $\xi = (\lambda, \rho)$. Based on $M_n = 5$, we obtain 51.51 as $T_{n,OS}$, which is much larger than the largest value of $T_{n,OS}$ in the 300 bootstrap samples (8.41) resulting in a p-value of zero, i.e., there is no support for the null hypothesis. Also here the choice of M_n is not crucial, the same p-value is obtained when setting $M_n = 3$. For this data set, the extended gamma frailty density with $m = 4$ was preferred: with $C_{0.05}$ estimated to be 3.93, $\text{AIC}_{C_{0.05}}(0)$ for the null model equals -30538.06 while for $m = 4$ $\text{AIC}_{C_{0.05}}(4)$ takes a value of -30435.46. The parameter estimates of the chosen density are $\widehat{\lambda} = 0.04 \cdot 10^{-2}$ (se $= 0.03 \cdot 10^{-3}$), $\widehat{\rho} = 1.76$ (se $= 0.01$), $\widehat{\beta} = -0.16$ (se $= 0.02$), $\widehat{\theta} = 6.02$ (se $= 5.32$),

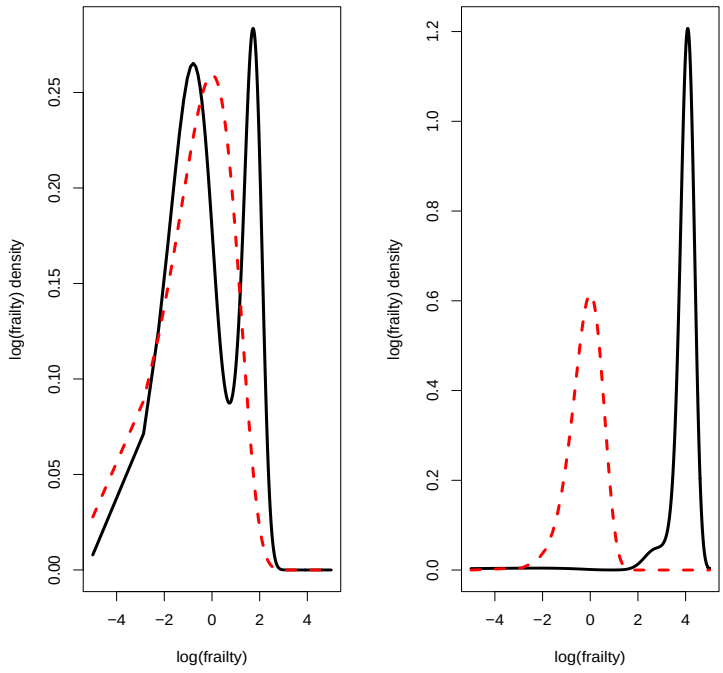


Figure 3.4: Estimated null density of $\log(U)$ (dashed red line) and estimated selected density of $\log(U)$ (solid black line). Left panel: udder infection data, right panel: insemination data.

$\widehat{d}_1 = 1.31$ (se $= 0.29$), $\widehat{d}_2 = 2.50$ (se $= 0.51$), $\widehat{d}_3 = 2.46$ (se $= 0.46$) and $\widehat{d}_4 = 1.77$ (se $= 0.56$). Figure 3.4 gives a graphical representation of the estimated null density and the estimated frailty density with $m = 4$.

The plot in the right panel of Figure 3.5 shows the centered average of the posterior mean frailty over time together with the 95% confidence limits (Shih and Louis, 1995) for the insemination data. Since the observed pattern is far from constant, the gamma frailty model does not fit the data well. Again this conclusion is in line with our finding.

3.6 Discussion

In this chapter, we unfold a procedure to assess the suitability of an one-parameter gamma density for the random effect in a shared frailty model with a parametric baseline

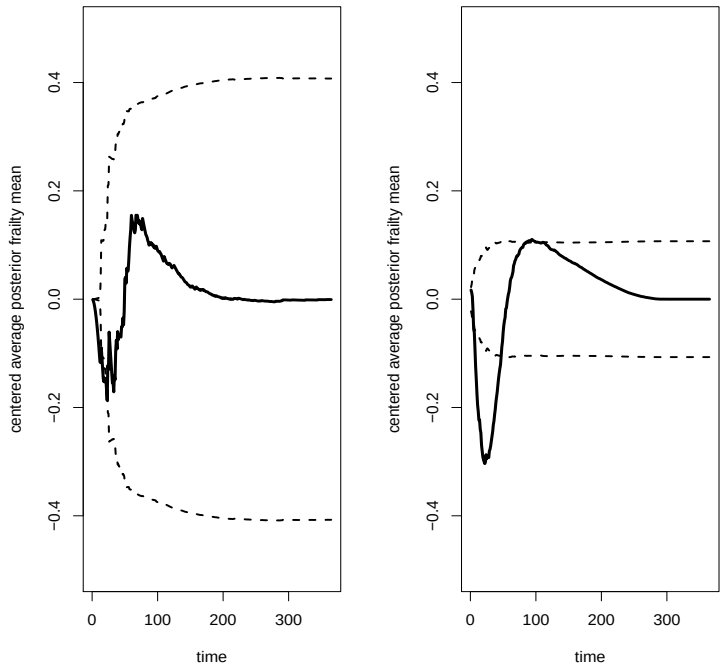


Figure 3.5: Centered average posterior frailty mean over time (Shih and Louis, 1995) - left panel: udder infection data, right panel: insemination data.

hazard. Using an orthonormal polynomial expansion, a new class of gamma frailty densities that enables us to approximate any continuous density on the positive half-line is defined. We obtain an explicit expression of the corresponding marginal likelihood for right-censored event time data. An order selection test based on an adapted version of Akaike's information criterion is used to find the best fitting model within a considered series of expanded densities. A bootstrap algorithm leads to an approximate p-value. It turns out that our approach tackles the problem at hand quite well, i.e., it is able to detect smaller as well as substantial deviations from an one-parameter gamma density with reasonable to very high power.

Even though the developed method proves to be useful, some remarks need to be made. First, the use of an order selection test requires the fit of multiple extended gamma frailty densities and hence is computationally intensive. A score test that relies only on the fit of the null density might serve as an alternative. For linear mixed models, such a score

test has been suggested by Thas (2009) in a discussion of Claeskens and Hart (2009). A score test would, however, require the calculation of the matrix of second order partial derivatives with respect to all unknown parameters in the shared frailty model, which is a non-trivial task. A singleton test might also be an alternative. Here, an extended gamma frailty density is formed by a single orthonormal polynomial, i.e., $\check{f}_{U_m}(u) = f_U(u)v_m^2(u)$ $(m = 0, \ldots, M_n)$. Note that, except for the null density, no nesting is present. A test based on $T_{n,\text{singleton}} = \max_{m=1,\ldots,M_n}\{\log \check{L}_m(\widehat{\xi}_m, \widehat{\beta}_m, \widehat{\theta}_m) - \log \check{L}_0(\widehat{\xi}_0, \widehat{\beta}_0, \widehat{\theta}_0)\}$, where $\check{L}_m$ is the likelihood of a shared frailty model with the mth singleton density $(m = 0, \ldots, M_n)$, can be performed. In a regression context, such a singleton test has been studied by Aerts *et al.* (2004). Similar to our test, a bootstrap procedure would be advised rather than working with the asymptotic test distribution.

Second, besides a random effect the shared frailty model includes two other components, namely the baseline hazard function and the exponential function describing the linear impact of one or more covariates. We focus on the frailty density, while the other (parametric) parts are presumed to be correct. This approach might be too restrictive, i.e., there is no guarantee that (for real-life data) a detection of lack-of-fit is not induced by a false specification of a non-frailty component. To allow for an unspecified baseline hazard function, several approaches may be applicable. One method is to profile out the baseline hazard via the EM algorithm for a shared frailty model (Duchateau and Janssen, 2008, section 5.1), whereas another strategy is to adjust the penalized likelihood algorithm for a shared frailty model (Duchateau and Janssen, 2008, section 5.2) to allow for an expanded gamma frailty density. Both approaches would use an iterative procedure to estimate the frailty density, while within the present parametric context an exact likelihood expression is available. To ease the linear impact of one or more covariates, the spline approach of Duchateau and Janssen (2004) might be worthwhile to investigate.

Third, the defined orthonormal polynomials are specifically constructed for an one-parameter gamma density. Similar calculations could be done for other frailty densities. While the derivation of the polynomials might still be explicit, there is no guarantee that the corresponding marginal likelihood has a closed form.

Part II

The copula model

Chapter 4

The copula model

In this chapter we briefly review some basic concepts of copula theory. Copulas can be used to unravel the association in grouped event time data if the clusters are small and of equal size. Denote the size of a group by p, i.e., for each cluster s ($s = 1, \ldots, n$) we have that $n_s = p$.

A p-variate copula is a function $C : [0,1]^p \to [0,1]$ satisfying:

(1) for every $(u_1, \ldots, u_p)$ in $[0,1]^p$: $C(u_1, \ldots, u_p) = 0$ if at least one $u_r = 0$ and $C(u_1, \ldots, u_p) = u_k$ if all $u_r = 1$ except u_k ($k \neq r = 1, \ldots, p$),

(2) for every hyperrectangle $B = [u_1, v_1] \times \ldots \times [u_p, v_p] \subset [0,1]^p$ with $u_r \leq v_r$ for all r ($r = 1, \ldots, p$), the C-volume of B is non-negative, i.e.,

$$\triangle_{u_p}^{v_p} \triangle_{u_{p-1}}^{v_{p-1}} \ldots \triangle_{u_2}^{v_2} \triangle_{u_1}^{v_1} C(w_1, \ldots, w_p) \geq 0$$

where $\triangle_{u_r}^{v_r} C(w_1, \ldots, w_p) = C(w_1, \ldots, v_r, \ldots, w_p) - C(w_1, \ldots, u_r, \ldots, w_p)$ ($r = 1, \ldots, p$).

Stated otherwise, a copula C is a p-variate distribution function on $[0,1]^p$ with uniform margins.

The key result in copula theory is the Sklar's theorem (Sklar, 1959): given a p-variate joint survival function $S(t_1, \ldots, t_p) = P(T_1 > t_1, \ldots, T_p > t_p)$ with marginal survival functions $S_r(t_r) = P(T_r > t_r)$ ($r = 1, \ldots, p$), there exists a p-variate copula C such that

$$S(t_1, \ldots, t_p) = C(S_1(t_1), \ldots, S_p(t_p)).$$

Moreover, if the marginal survival functions are continuous, then the copula C is unique and

$$C(u_1, \ldots, u_p) = S(S_1^{-1}(u_1), \ldots, S_p^{-1}(u_p))$$

with S_r^{-1} quantile functions $(r = 1, \ldots, p)$.

The above implies that the modeling of the joint survival function can be split into two distinct parts: (1) the specification of the marginal survival functions; (2) the identification of a copula that interconnects the former and hence fully captures the within-cluster association. Note that such a complete segregation is not possible for a shared frailty model (Goethals *et al.*, 2008).

Often the pairwise correlation of cluster items is of particular interest. The strength of the association modeled by a bivariate copula C is commonly expressed via Kendall's tau (τ), which is defined as (Nelsen, 2006)

$$\tau \quad = \quad 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

It holds that $\tau \in [-1, 1]$ where independence corresponds to $\tau = 0$. Since Kendall's tau is a global measure of association, it does not reflect the tail behavior of the data. To investigate the latter, the lower (δ_L) and upper (δ_U) tail dependence coefficients can be used. With U_1 and U_2 uniform $[0, 1]$ random variables, these measures are given by (Nelsen, 2006)

$$\begin{aligned} \delta_L &= \lim_{u \to 0} P(U_2 < u | U_1 < u) = \lim_{u \to 0} \frac{C(u, u)}{u} \\ \delta_U &= \lim_{u \to 1} P(U_2 > u | U_1 > u) = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u}. \end{aligned}$$

C is said to have lower (upper) tail dependence if $\delta_L \in]0, 1]$ $(\delta_U \in]0, 1])$ and no lower (upper) tail dependence if $\delta_L = 0$ $(\delta_U = 0)$. Here, a copula is used to model the joint survival function of event times. Therefore, a copula with lower tail dependence covers the association between late event times, whereas a copula with upper tail dependence captures the association between events that occur early in time.

In this dissertation, the focus is on semiparametric (Chapter 5) and nonparametric (Chapter 6) copula estimation. In a semiparametric copula, the marginal survival functions are modeled in a nonparametric way, while the copula function is taken to be of a parametric form. On the other hand, if the marginal survival functions and the copula

function are left unspecified, we obtain a nonparametric copula. Often, a semiparametric copula is constructed via one or several Laplace transforms. With M the distribution function of a positive random variable, a Laplace transform is defined as

$$\psi(s) = \int_0^\infty \exp(-s\alpha) dM(\alpha)$$

or equivalently, a Laplace transform is a continuous strictly decreasing function $\psi = [0, \infty[\to [0, 1]$ with $\psi(0) = 1$, $\psi(\infty) = 0$ and satisfying the complete monotonicity condition $(-1)^k \frac{\partial^k}{\partial s^k} \psi(s) \geq 0$, i.e., its derivatives must alternate in sign.

More (basic) information on copulas can be found in Nelsen (2006).

Chapter 5

Semiparametric copula estimation

As mentioned in Chapter 4, a copula can be used to model clustered time-to-event data. In this chapter, the focus is on four-variate semiparametric copulas, i.e., the marginal survival functions are modeled in a nonparametric way while the copula function takes a parametric form.

Popular models are the exchangeable and the nested Archimedean copulas (Nelsen, 2006; Savu and Trede, 2006; Hofert, 2008). Unfortunately, they induce rather restrictive data dependence structures. Therefore, we investigate the use of Joe-Hu copulas (Joe and Hu, 1996) as a more flexible alternative. To fit the diverse copulas, we use the likelihood approach by Shih and Louis (1995). Given the variety of possible copulas, the question of model selection arises. Based on ideas in Sin and White (1996), Claeskens and Hjort (2008) as well as in Chen *et al.* (2010), we state conditions under which a penalized likelihood based information criterion is weakly consistent or consistent.

In Section 5.1 the exchangeable and the nested Archimedean copulas are compared to the more flexible Joe-Hu copulas. Section 5.2 contains the quasi-likelihood for right-censored quadruple event time data. In Section 5.3 we address the issue of model selection. In Section 5.4 the udder infection data are analyzed. To investigate the numerical performance of the methodology we set up a simulation study, see Section 5.5. Conclusions and ideas for future work are in Section 5.6. Supplementary info on the tail behavior of exchangeable Archimedean and Joe-Hu copulas is given in Section 5.7. Key

results needed for the proofs are collected in Appendix B.1.

5.1 Flexible copula models

In this section we show that the association pattern induced by the exchangeable and the nested Archimedean copulas is quite restrictive and that the alternative Joe-Hu copulas provide more flexibility.

5.1.1 Exchangeable and nested Archimedean copulas

With ψ a Laplace transform, a four-variate exchangeable Archimedean copula (EAC) is defined as

$$C(u_1, \ldots, u_4) = \psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \psi^{-1}(u_3) + \psi^{-1}(u_4)\right) \tag{5.1}$$

and is therefore completely determined by the choice of ψ. The latter implies a restrictive dependence structure. Indeed, since only ψ can be specified, all bivariate marginal copulas are exactly the same. For EAC, ψ is often called the generator of the copula.

Fully and partially nested Archimedean copulas (FNAC, resp. PNAC) extend EAC's and hence allow a more flexible association pattern. For four-dimensional data, the correlation structures are given in Figure 5.1. The corresponding copula expressions are

$$\begin{aligned} &C(u_1, \ldots, u_4) \\ &\quad = \psi_3\left(\psi_3^{-1}(u_4) + \psi_3^{-1}\left[\psi_2\left(\psi_2^{-1}(u_3) + \psi_2^{-1}\left[\psi_1\left(\psi_1^{-1}(u_2) + \psi_1^{-1}(u_1)\right)\right]\right)\right]\right) \end{aligned} \tag{5.2}$$

respectively,

$$\begin{aligned} &C(u_1, \ldots, u_4) \\ &\quad = \psi_3\left(\psi_3^{-1}\left[\psi_1\left(\psi_1^{-1}(u_1) + \psi_1^{-1}(u_2)\right)\right] + \psi_3^{-1}\left[\psi_2\left(\psi_2^{-1}(u_3) + \psi_2^{-1}(u_4)\right)\right]\right) \end{aligned} \tag{5.3}$$

where ψ_i $(i = 1, 2, 3)$ are Laplace transforms. From the formulas it follows that FNAC and PNAC allow the free specification of three out of the six bivariate margins, while the three remaining bivariate margins are implied by the chosen copula structure. Further, to be a valid copula, each combination of Laplace transforms within a NAC, as given by $\psi_i^{-1} \circ \psi_j$ $(i \neq j \in \{1, 2, 3\})$, needs to satisfy the complete monotonicity condition (Joe, 1997; Hofert, 2008). For two Laplace transforms from the same family the latter is equivalent to claiming that the degree of dependence, as expressed by the bivariate copula parameters, decreases with the level of nesting. However, the mixing of diverse Laplace transforms needs to be handled with more care. We illustrate the complete monotonicity condition for a NAC with two examples.

Example 1. Consider a Gumbel Laplace transform ψ with parameter θ_0 and another Gumbel Laplace transform ϕ with parameter θ_1. It holds that

$$\psi^{-1} \circ \phi(s) = s^{\frac{\theta_0}{\theta_1}}.$$

The first order derivative is always positive. The second order derivative is negative if and only if $\theta_0 \leq \theta_1$. Given the latter, the alternating signs of the higher order derivatives is ensured. Therefore $\psi^{-1} \circ \phi$ satisfies the complete monotonicity condition if and only if $\theta_0 \leq \theta_1$.

Example 2. Consider a Clayton Laplace transform ψ with parameter θ_0 and a Gumbel Laplace transform ϕ with parameter θ_1. It holds that

$$\psi^{-1} \circ \phi(s) = \frac{\exp\left(\theta_0 s^{\frac{1}{\theta_1}}\right) - 1}{\theta_0}.$$

The first order derivative is always positive. The second order derivative is only negative for s-values satisfying

$$\left(\frac{\theta_1 - 1}{\theta_0}\right)^{\theta_1} \geq s$$

regardless the value of θ_0 and θ_1. Therefore $\psi^{-1} \circ \phi$ does not satisfy the complete monotonicity condition, i.e., the combination of a Clayton Laplace transform with a Gumbel Laplace transform will always leads to an improper copula.

5.1.2 Joe-Hu copulas

A copula family that is much more flexible than the one of exchangeable and nested Archimedean copulas is the Joe-Hu family (Joe and Hu, 1996), which is constructed as follows. Let K_{ij} be bivariate copulas that are max-id, i.e., K_{ij}^{γ} is a distribution function for all $\gamma > 0$ $(1 \leq i < j \leq 4)$. Further, let $H_1, \ldots, H_4$ be univariate cdf's on $[0, 1]$ and let M be the distribution function of a positive random variable α whose Laplace transform is given by ψ. Joe and Hu (1996) define a four-variate copula by considering the following mixture:

$$\begin{aligned}
&\int_0^\infty \prod_{1 \leq i < j \leq 4} K_{ij}^{\alpha}(H_i(u_i), H_j(u_j)) \prod_{i=1}^{4} H_i^{\nu_i \alpha}(u_i)\, dM(\alpha) \\
&= \psi\Big(-\log\Big(\prod_{1 \leq i < j \leq 4} K_{ij}(H_i(u_i), H_j(u_j)) \prod_{i=1}^{4} H_i^{\nu_i}(u_i) \Big) \Big) \\
&= \psi\Big(-\sum_{1 \leq i < j \leq 4} \log K_{ij}(H_i(u_i), H_j(u_j)) - \sum_{i=1}^{4} \nu_i \log H_i(u_i) \Big) \qquad (5.4)
\end{aligned}$$

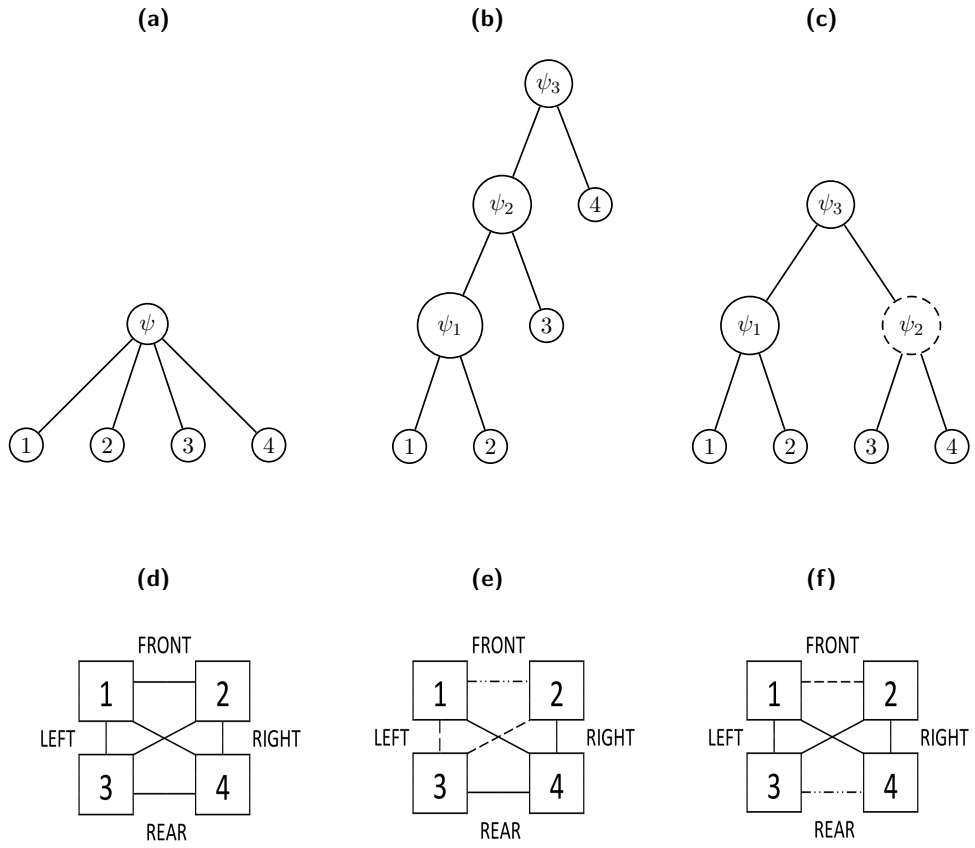


Figure 5.1: With θ the parameter of ψ and θ_j the parameter of ψ_j $(j = 1, 2, 3)$, the tree structures and the association patterns of EAC, FNAC and PNAC are given by: (a) and (d) EAC: all pairs have the same dependence parameter θ; (b) and (e) FNAC: pair (1,2) has association parameter θ_1, pairs (1,3) and (2,3) have association parameter θ_2 and all other pairs have association parameter θ_3 (with $\theta_1 \geq \theta_2 \geq \theta_3$); (c) and (f) PNAC: pair (1,2), resp. pair (3,4), has association parameter θ_1, resp. θ_2, and all other pairs have association parameter θ_3 (with $\theta_1 \geq \theta_3, \theta_2 \geq \theta_3$). A bigger node corresponds to a larger value for θ_j, while different line types represent different dependencies.

where the ν_i's are chosen fixed constants satisfying $\nu_i > -3$ $(i = 1, \ldots, 4)$ (for dimension p: $\nu_i > 1 - p$). The ν_i's are usually nonnegative, but they can be negative if some of the K_{ij} correspond to an independence copula.

For (5.4) to be a copula, the margins need to be uniform. This can be achieved by choosing $H_i(u_i)$ equal to $\exp(-p_i\psi^{-1}(u_i))$ with $p_i = 1/(\nu_i + 3)$ $(i = 1, \ldots, 4)$ (for dimension p: $p_i = 1/(\nu_i + p - 1)$). One then obtains the copula

$$\begin{aligned} & C(u_1, \ldots, u_4) \\ &= \psi\Big(- \sum_{1 \le i < j \le 4} \log K_{ij}(\exp(-p_i\psi^{-1}(u_i)), \exp(-p_j\psi^{-1}(u_j))) + \sum_{i=1}^{4} \nu_i p_i \psi^{-1}(u_i) \Big). \end{aligned} \tag{5.5}$$

Note that the inclusion of the ν_i's ensures that the above family of multivariate copulas is closed under margins and that the (i,j)-bivariate marginal copula is given by

$$\begin{aligned} C(u_i, u_j) &= \psi(-\log K_{ij}(\exp(-p_i\psi^{-1}(u_i)), \exp(-p_j\psi^{-1}(u_j))) \\ &\quad + (\nu_i + 2)p_i\psi^{-1}(u_i) + (\nu_j + 2)p_j\psi^{-1}(u_j)). \end{aligned} \tag{5.6}$$

See Joe and Hu (1996) for technical details where it is further shown that $C(u_i, u_j)$ is more concordant, i.e., more positive quadrant dependent than its Archimedean counterpart.

The dependence structure in (5.5) is completely determined by the choice of the K_{ij} $(1 \le i < j \le 4)$ and the Laplace transform ψ. The Laplace transform ψ induces a minimal level of overall association, while the copulas K_{ij} allow a fine-tuning of the dependence for each of the six bivariate margins. Examples of popular bivariate Archimedean copulas that are max-id and therefore can be used as building blocks K_{ij} are listed in Table 5.1. These copulas have different dependence properties, hence using (a combination of) them in (5.5) allows the construction of copulas with flexible dependence patterns. We refer to Section 5.7 for a description and a visualization of the dependence properties of bivariate Archimedean copulas and bivariate Joe-Hu models. The impact of the chosen ν_i on the modeled association is also illustrated. Conditions to check the max-id assumption can be found in Joe (1993) and Joe and Hu (1996).

Note that by taking $K_{ij}(u_i, u_j) = u_i u_j$, i.e., the independence copula, for all pairs (i, j) $(i \ne j \in \{1, \ldots, 4\})$, the copula in (5.5) is Archimedean with Laplace transform ψ. Furthermore, by appropriate choices of the K_{ij} $(1 \le i < j \le 4)$ with $K_{ij}(u_i, u_j) \ne u_i u_j$ for some pairs (i, j) it is possible to create a dependence structure that is the same as the

Table 5.1: Bivariate Archimedean max-id copulas and their Laplace transforms - $\tilde{u} = 1 - u$ and $\tilde{v} = 1 - v$.

copula	$K(u,v)$	$\psi(s)$	$\theta \in$
Clayton	$\left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$	$(1+\theta s)^{-1/\theta}$	$]0,\infty[$
Gumbel	$e^{-\left((-\log u)^{\theta} + (-\log v)^{\theta}\right)^{1/\theta}}$	$e^{-s^{1/\theta}}$	$[1,\infty[$
Frank	$-\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$-\frac{1}{\theta}\log\left(1-(1-e^{-\theta})e^{-s}\right)$	$]0,\infty[$
Joe	$1 - \left(\tilde{u}^{\theta} + \tilde{v}^{\theta} - \tilde{u}^{\theta}\tilde{v}^{\theta}\right)^{1/\theta}$	$1-(1-e^{-s})^{1/\theta}$	$[1,\infty[$

one generated by a nested Archimedean copula and this without the required modeling restrictions of the latter, e.g., the association structure in Figure 5.1 (e) can be obtained by taking $K_{12} \neq K_{13} = K_{23} \neq K_{14} = K_{24} = K_{34}$, while the correlation pattern in Figure 5.1 (f) can be constructed by setting $K_{12} \neq K_{34} \neq K_{13} = K_{23} = K_{14} = K_{24}$. Exchangeable and nested Archimedean copulas are thus, in some sense, a subclass of the Joe-Hu family.

5.2 The quasi-loglikelihood

Consider four-variate right-censored event time data. Denote the observed time of item r $(r = 1,\ldots,4)$ in cluster s $(s = 1,\ldots,n)$ by $Y_{sr} = \min(T_{sr}, C_{sr})$ with T_{sr} the true event time and C_{sr} the censoring time. The indicator $\delta_{sr} = I(T_{sr} \leq C_{sr})$ equals one if $Y_{sr} = T_{sr}$ and zero otherwise. Event times and censoring times are assumed to be independent. Denote the joint survival function of $(T_1,\ldots,T_4)$ by S and the marginal survival functions by S_r $(r = 1,\ldots,4)$. Recall from Sklar's theorem that $S(t_1,\ldots,t_4) = C(S_1(t_1),\ldots,S_4(t_4))$ for a copula function C. We aim at modeling C in a semiparametric way.

Consider a set $\mathcal{M}$ of diverse EAC, FNAC and PNAC as well as Joe-Hu models:

$$\mathcal{M} = \bigcup_{d=1}^{D} M_d$$

with $M_d = \{C_d(u_1,\ldots,u_4;\zeta_d) : \zeta_d \in A_d \subset \mathbb{R}^{q_d}\}$. Here, the parameter space A_d is a subset of the q_d-dimensional Euclidean space. The parameter vector ζ_d contains, for each single model M_d, all specific parametric characteristics: the Laplace transform ψ_d (EAC, Joe-Hu), the Laplace transforms $\psi_{d,i}$ (FNAC, PNAC) $(i = 1,2,3)$ and the bivariate max-id copulas $K_{d,ij}$ (Joe-Hu) $(1 \leq i < j \leq 4)$.

Further, with

$$\begin{aligned}
\Delta_s &= \prod_{r=1}^{4}(1-\delta_{sr}) \\
\Delta_s(p) &= \delta_{sp}\prod_{r=1;r\neq p}^{4}(1-\delta_{sr}) \\
\Delta_s(p,q) &= \delta_{sp}\delta_{sq}\prod_{r=1;r\neq p,q}^{4}(1-\delta_{sr}) \quad \text{for } p\neq q \\
\Delta_s(p,q,v) &= \delta_{sp}\delta_{sq}\delta_{sv}(1-\delta_{sw}) \quad \text{for } w\neq p,q,v \text{ and } p\neq q\neq v \\
\Delta_s(1,2,3,4) &= \prod_{r=1}^{4}\delta_{sr}
\end{aligned}$$

and

$$\begin{aligned}
&l_{s,d}(u_{s1},\ldots,u_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta_d) \\
&\quad= \Delta_s\log\left(C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)\right) \\
&\qquad+\sum_{p=1}^{4}\left[\Delta_s(p)\log\left(\frac{\partial C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{sp}}\right)\right] \\
&\qquad+\sum_{p\neq q}\left[\Delta_s(p,q)\log\left(\frac{\partial^2 C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{sp}\partial u_{sq}}\right)\right] \\
&\qquad+\sum_{p\neq q\neq v}\left[\Delta_s(p,q,v)\log\left(\frac{\partial^3 C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{sp}\partial u_{sq}\partial u_{sv}}\right)\right] \\
&\qquad+\Delta_s(1,2,3,4)\log\left(\frac{\partial^4 C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{s1}\partial u_{s2}\partial u_{s3}\partial u_{s4}}\right)
\end{aligned}$$

the loglikelihood of model M_d ($d=1,\ldots,D$) for right-censored data is given by (Shih and Louis, 1995; Massonnet *et al.*, 2009)

$$\log L_{n,d}(\zeta_d)=\sum_{s=1}^{n} l_{s,d}(V_{s1},\ldots,V_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta_d)$$

where $V_{sr}=S_r(Y_{sr})$ ($s=1,\ldots,n$ and $r=1,\ldots,4$). From the above ζ_d can be estimated if S_r ($r=1,\ldots,4$) are known. Often S_r ($r=1,\ldots,4$) are unknown and the two-step approach by Shih and Louis (1995) is applied: replace S_r by its Kaplan-Meier counterpart $\widehat{S}_r$ to obtain $\widehat{V}_{sr}=\widehat{S}_r(Y_{sr})$ ($s=1,\ldots,n$ and $r=1,\ldots,4$) and maximize the quasi-loglikelihood

$$\log \widetilde{L}_{n,d}(\zeta_d)=\sum_{s=1}^{n} l_{s,d}(\widehat{V}_{s1},\ldots,\widehat{V}_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta_d)$$

to attain the maximum quasi-likelihood estimator

$$\widehat{\zeta}_{n,d} = \arg \max_{\zeta_d \in A_d} \log \widetilde{L}_{n,d}(\zeta_d).$$

The optimization of the quasi-loglikelihood is implemented in a generic R-program which is discussed in Section 5.5.

5.3 Model selection

We would like to select a model from the set $\mathcal{M} = \bigcup_{d=1}^{D} M_d$. Misspecification is possible, i.e., the true (unknown) copula C does not need to be in $\mathcal{M}$. Hence, we want to choose the 'best possible' copula in $\mathcal{M}$ or the copula in $\mathcal{M}$ that is the 'closest' we can get to the true (unknown) copula C (Sin and White, 1996; Claeskens and Hjort, 2008). The discrepancy between the true (unknown) copula C and model M_d in $\mathcal{M}$ $(d = 1, \ldots, D)$ can be measured by the Kullback-Leibler discrepancy

$$n^{-1}\mathsf{E}_{\text{true}}\left[\log L_{\text{true}}(\zeta_{\text{true}}) - \log L_{n,d}(\zeta_d)\right] \tag{5.7}$$

where E_{true} denotes the expectation with respect to the true (unknown) copula C and $\log L_{\text{true}}$ is the true (unknown) loglikelihood.

Given that we work under general misspecification, we need, for each model M_d $(d = 1, \ldots, D)$, the pseudo-true parameter value $\zeta^*_{n,d}$ where

$$\zeta^*_{n,d} = \arg \max_{\zeta_d \in A_d} n^{-1} \sum_{s=1}^{n} \mathsf{E}_{\text{true}}\left[l_{s,d}(V_1, \ldots, V_4, \delta_1, \ldots, \delta_4; \zeta_d)\right].$$

The pseudo-true parameter value minimizes the Kullback-Leibler discrepancy (5.7) or equivalently, it is the parameter value for which the loglikelihood using the dth model is as close as possible to the true (unknown) loglikelihood. Note that for the data generating process that we consider, i.e., right-censored data with independent event and censoring times, the above reduces to

$$\zeta^*_{n,d} = \zeta^*_d = \arg \max_{\zeta_d \in A_d} \mathsf{E}_{\text{true}}\left[l_d(V_1, \ldots, V_4, \delta_1, \ldots, \delta_4; \zeta_d)\right].$$

In the remainder we present the model selection procedure in its general form.

To select a model from the set $\mathcal{M} = \bigcup_{d=1}^{D} M_d$ a penalized quasi-loglikelihood is used. More specifically, for model M_d $(d = 1, \ldots, D)$ we consider

$$\mathsf{IC}_n(M_d) = -2 \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) + \mathsf{pen}(n, d). \tag{5.8}$$

Common examples of the penalty term are $\text{pen}(n,d) = 2q_d$, leading to the Akaike information criterion, AIC (Akaike, 1973) and $\text{pen}(n,d) = q_d \log n$, resulting in the Bayesian information criterion, BIC (Schwarz, 1978). The smaller the value of $\text{IC}_n(M_d)$, the better the model M_d is.

Propositions 5.3.1 and 5.3.2 indicate that, based on information criterion (5.8), the best model is chosen in a (weakly) consistent way. Here, we follow the terminology introduced by Sin and White (1996), where (weak) consistency of a selection criterion refers to the use of a weak law of large numbers. The proofs of Propositions 5.3.1 and 5.3.2 rely on results in Chen *et al.* (2010), where it is shown that under suitable conditions: (1) the estimator $\widehat{\zeta}_{n,d}$ converges to the pseudo-true parameter value $\zeta^*_{n,d}$ and (2) the estimator $\widehat{\zeta}_{n,d}$ is asymptotically normal. The findings of Chen *et al.* (2010) and the corresponding conditions are listed in Appendix B.1.

Suppose that there is exactly one model M_{d_0} in the set of considered models $\mathcal{M}$ which reaches the smallest Kullback-Leibler discrepancy. Proposition 5.3.1 states a condition on the penalty under which information criterion (5.8) is weakly consistent, i.e., under which it selects model M_{d_0} with probability tending to one.

Proposition 5.3.1 (Weak consistency)**.** *Suppose that there is an unique model M_{d_0} in the set of considered models $\mathcal{M}$ which reaches the smallest Kullback-Leibler discrepancy, i.e.,*

$$\liminf_{n\to\infty} \min_{d\neq d_0} n^{-1} \mathsf{E}_{\text{true}}[\log L_{n,d_0}(\zeta^*_{n,d_0}) - \log L_{n,d}(\zeta^*_{n,d})] > 0. \tag{5.9}$$

Define $\Delta\text{IC}_n(d_0,d) = \text{IC}_n(M_{d_0}) - \text{IC}_n(M_d)$. If for all $d = 1,\ldots,D$, $\text{pen}(n,d_0) - \text{pen}(n,d) = o_p(n)$ and if the conditions of Proposition B.1.1 are satisfied, then weak consistency holds, i.e.,

$$\lim_{n\to\infty} P(\max_{d\neq d_0} \Delta\text{IC}_n(d_0,d) < 0) = 1.$$

Proof. From Proposition B.1.1 we have ($d_0 \neq d = 1,\ldots,D$)

$$n^{-1}\{\log \widetilde{L}_{n,d_0}(\widehat{\zeta}_{n,d_0}) - \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d})\} \\ = n^{-1}\mathsf{E}_{\text{true}}[\log L_{n,d_0}(\zeta^*_{n,d_0}) - \log L_{n,d}(\zeta^*_{n,d})] + Q_n$$

with $Q_n = o_P(1)$ or equivalently

$$\log \widetilde{L}_{n,d_0}(\widehat{\zeta}_{n,d_0}) - \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) = \mathsf{E}_{\text{true}}[\log L_{n,d_0}(\zeta^*_{n,d_0}) - \log L_{n,d}(\zeta^*_{n,d})] + Q^*_n$$

with $Q_n^* = o_P(n)$. Hence, for models M_{d_0} and M_d satisfying (5.9) ($d_0 \neq d = 1, \ldots, D$) we obtain

$$\begin{aligned}
\Delta\mathsf{IC}_n(d_0, d) &= \mathsf{IC}_n(M_{d_0}) - \mathsf{IC}_n(M_d) \\
&= -2\{\log \widetilde{L}_{n,d_0}(\widehat{\zeta}_{n,d_0}) - \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d})\} + \mathsf{pen}(n, d_0) - \mathsf{pen}(n, d) \\
&= -2\{\mathsf{E}_{\text{true}}[\log L_{n,d_0}(\zeta^*_{n,d_0}) - \log L_{n,d}(\zeta^*_{n,d})]\} + Q_n^* \\
&\quad +\mathsf{pen}(n, d_0) - \mathsf{pen}(n, d) \\
&= -2\{\mathsf{E}_{\text{true}}[\log L_{n,d_0}(\zeta^*_{n,d_0}) - \log L_{n,d}(\zeta^*_{n,d})]\} + R_n \\
&\leq -\kappa n + R_n
\end{aligned}$$

with $R_n = Q_n^* + \mathsf{pen}(n, d_0) - \mathsf{pen}(n, d)$ and where, for $\kappa > 0$, the last equality follows from (5.9).

The former holds for all models M_{d_0} and M_d satisfying (5.9) ($d_0 \neq d = 1, \ldots, D$), we thus have $\max_{d \neq d_0} \Delta\mathsf{IC}_n(d_0, d) \leq -\kappa n + R_n$.

Further, due to the penalty condition, it holds that for each $\xi > 0$ there exists $N(\xi)$ such that for $n > N(\xi)$: $P(-\kappa n + R_n > 0) < \xi$ or $P(-\kappa n + R_n < 0) > 1 - \xi$.

Consequently, $\lim_{n\to\infty} P(\max_{d \neq d_0} \Delta\mathsf{IC}_n(d_0, d) < 0) = 1$. □

Proposition 5.3.1 implies that model selection based on AIC and/or BIC is weakly consistent.

Next, suppose that there are two or more models that achieve nearly the same small Kullback-Leibler discrepancy. Define $\mathcal{J}$ and $\widetilde{\mathcal{J}}$ to be (non-empty) sets of such 'good' models and state that M_d, $M_{d'} \in \mathcal{J}$ if and only if

$$\limsup_{n\to\infty} n^{-1/2} \mathsf{E}_{\text{true}}[\log L_{n,d'}(\zeta^*_{n,d'}) - \log L_{n,d}(\zeta^*_{n,d})] < \infty \tag{5.10}$$

and M_d, $M_{d'} \in \widetilde{\mathcal{J}}$ if and only if

$$\log \widetilde{L}_{n,d'}(\zeta^*_{n,d'}) - \log \widetilde{L}_{n,d}(\zeta^*_{n,d}) = O_P(1). \tag{5.11}$$

Proposition 5.3.2 states conditions on the penalty under which information criterion (5.8) selects, with probability tending to one, from a set of 'good' models $\mathcal{J}$ or $\widetilde{\mathcal{J}}$ a most parsimonious model, i.e., a model having the least parameters. The latter is called consistency. Note that there might be more than one parsimonious model. Define $\mathcal{J}_0 \subset \mathcal{J}$ the subset of $\mathcal{J}$ with the most parsimonious models, i.e., $\mathcal{J}_0 = \{M_{d_0} \in \mathcal{J} : q_{d_0} =$

$\min\{q_d : M_d \in \mathcal{J}\}\}$ and $\widetilde{\mathcal{J}}_0 \subset \widetilde{\mathcal{J}}$ the subset of $\widetilde{\mathcal{J}}$ with the most parsimonious models, i.e., $\widetilde{\mathcal{J}}_0 = \{M_{d_0} \in \widetilde{\mathcal{J}} : q_{d_0} = \min\{q_d : M_d \in \widetilde{\mathcal{J}}\}\}$.

Proposition 5.3.2 (Consistency)**.** *Assume that the conditions of Proposition B.1.3 are satisfied.*

(a) With $\mathcal{J}$ defined in (5.10), assume that for all $M_{d_0} \in \mathcal{J}_0$, for all $M_d \in \mathcal{J}\backslash\mathcal{J}_0$, the penalty is such that $P(\{\mathit{pen}(n,d) - \mathit{pen}(n,d_0)\}/\sqrt{n} \to \infty) = 1$. Then, with probability tending to one a most parsimonious model will be selected, i.e.,

$$\lim_{n\to\infty} P(\max_{M_d \notin \mathcal{J}_0} \Delta \mathit{IC}_n(d_0, d) < 0) = 1.$$

(b) With $\widetilde{\mathcal{J}}$ defined in (5.11), assume that for all $M_{d_0} \in \widetilde{\mathcal{J}}_0$, for all $M_d \in \widetilde{\mathcal{J}}\backslash\widetilde{\mathcal{J}}_0$: $P(\mathit{pen}(n,d) - \mathit{pen}(n,d_0) \to \infty) = 1$. Then, with probability tending to one a most parsimonious model will be selected, i.e.,

$$\lim_{n\to\infty} P(\max_{M_d \notin \widetilde{\mathcal{J}}_0} \Delta \mathit{IC}_n(d_0, d) < 0) = 1.$$

Proof. We first prove (a).

From Proposition B.1.3 we have ($d_0 \neq d = 1, \ldots, D$)

$$\log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) - \log \widetilde{L}_{n,d_0}(\widehat{\zeta}_{n,d_0}) = \mathsf{E}_{\text{true}}[\log L_{n,d}(\zeta^*_{n,d}) - \log L_{n,d_0}(\zeta^*_{n,d_0})] + Q_n$$

with $Q_n = O_P(n^{1/2})$. Hence, for models M_{d_0} and M_d ($d_0 \neq d = 1, \ldots, D$), we obtain

$$\begin{aligned}
\Delta\mathsf{IC}_n(d_0, d) &= \mathsf{IC}_n(M_{d_0}) - \mathsf{IC}_n(M_d) \\
&= 2\{\log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) - \log \widetilde{L}_{n,d_0}(\widehat{\zeta}_{n,d_0})\} - \{\mathsf{pen}(n,d) - \mathsf{pen}(n,d_0)\} \\
&= 2\{\mathsf{E}_{\text{true}}[\log L_{n,d}(\zeta^*_{n,d}) - \log L_{n,d_0}(\zeta^*_{n,d_0})]\} + Q_n \\
&\quad -\{\mathsf{pen}(n,d) - \mathsf{pen}(n,d_0)\} \\
&= R_n - P_n
\end{aligned}$$

where $R_n = 2\{\mathsf{E}_{\text{true}}[\log L_{n,d}(\zeta^*_{n,d}) - \log L_{n,d_0}(\zeta^*_{n,d_0})]\} + Q_n$ and $P_n = \mathsf{pen}(n,d) - \mathsf{pen}(n,d_0)$.

The former holds for all models M_{d_0} and M_d ($d_0 \neq d = 1, \ldots, D$), thus also for all models $M_{d_0} \in \mathcal{J}_0$ and $M_d \in \mathcal{J}\backslash\mathcal{J}_0$. We then have

$$\max_{M_d \notin \mathcal{J}_0} \Delta\mathsf{IC}_n(d_0, d) = R^*_n - P^*_n$$

where R_n^* and P_n^* are R_n resp. P_n corresponding to $M_{d_0} \in \mathcal{J}_0$ and $M_d \in \mathcal{J}\backslash\mathcal{J}_0$ for which $\Delta\mathsf{IC}_n(d_0, d)$ is maximal.

Further, (5.10) implies that for all $\xi > 0$ there exists $\Delta(\xi) < \infty$ and $N_1(\xi)$ such that for all $n > N_1(\xi)$: $P(R_n^* > n^{1/2}\Delta(\xi)) < \xi/2$. Also, due to the penalty condition, there exists $N_2(\xi)$ such that for all $n > N_2(\xi)$: $P(P_n^* < n^{1/2}\Delta(\xi)) < \xi/2$.

Let $n > \max(N_1(\xi), N_2(\xi))$, then

$$\begin{aligned} P(\max_{M_d \notin \mathcal{J}_0} \Delta\mathsf{IC}_n(d_0, d) < 0) &= P(R_n^* < P_n^*) \\ &\geq P(R_n^* \leq n^{1/2}\Delta(\xi), P_n^* \geq n^{1/2}\Delta(\xi)) \\ &\geq 1 - P(R_n^* < n^{1/2}\Delta(\xi)) - P(P_n^* < n^{1/2}\Delta(\xi)) \\ &> 1 - \xi. \end{aligned}$$

Consequently, $\lim_{n\to\infty} P(\max_{M_d \notin \mathcal{J}_0} \Delta\mathsf{IC}_n(d_0, d) < 0) = 1$.

We now prove (b).

Define

$$\begin{aligned} \nabla \log L_{n,d}(\zeta_d) &= \frac{\partial \log L_{n,d}(\zeta_d)}{\partial \zeta_d}, \nabla^2 \log L_{n,d}(\zeta_d) = \frac{\partial^2 \log L_{n,d}(\zeta_d)}{\partial \zeta_d^2} \\ \nabla \log \widetilde{L}_{n,d}(\zeta_d) &= \frac{\partial \log \widetilde{L}_{n,d}(\zeta_d)}{\partial \zeta_d}, \nabla^2 \log \widetilde{L}_{n,d}(\zeta_d) = \frac{\partial^2 \log \widetilde{L}_{n,d}(\zeta_d)}{\partial \zeta_d^2}. \end{aligned}$$

A Taylor expansion of $\log \widetilde{L}_{n,d}(\zeta_{n,d}^*)$ around $\widehat{\zeta}_{n,d}$ gives

$$\begin{aligned} \log \widetilde{L}_{n,d}(\zeta_{n,d}^*) &= \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) + (\zeta_{n,d}^* - \widehat{\zeta}_{n,d})\, \nabla \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) \\ &+ \frac{1}{2}(\zeta_{n,d}^* - \widehat{\zeta}_{n,d})^T\, \nabla^2 \log \widetilde{L}_{n,d}(\bar{\zeta}_{n,d})\, (\zeta_{n,d}^* - \widehat{\zeta}_{n,d}) \\ &= \log \widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) + \frac{1}{2}(\zeta_{n,d}^* - \widehat{\zeta}_{n,d})^T \nabla^2 \log \widetilde{L}_{n,d}(\bar{\zeta}_{n,d})\, (\zeta_{n,d}^* - \widehat{\zeta}_{n,d}) \end{aligned}$$

where $\bar{\zeta}_{n,d}$ is between $\zeta_{n,d}^*$ and $\widehat{\zeta}_{n,d}$.

Define $B_{n,d} = -n^{-1}\mathsf{E}_{\text{true}}[\nabla^2 \log L_{n,d}(\zeta_{n,d}^*)]$ and $\bar{B}_{n,d} = -n^{-1}\nabla^2 \log \widetilde{L}_{n,d}(\bar{\zeta}_{n,d})$. Under the conditions of Proposition B.1.2, Chen et al. (2010) show that $\bar{B}_{n,d}$ converges in probability to $B_{n,d}$ as $n \to \infty$.

Further,

$$\begin{aligned}&\frac{1}{2}(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})^T\,\bar{B}_{n,d}\,(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})\\ &\quad=\quad\frac{1}{2}(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})^T\,(\bar{B}_{n,d}-B_{n,d})\,(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})+\frac{1}{2}(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})^T\,B_{n,d}\,(\zeta^*_{n,d}-\widehat{\zeta}_{n,d}).\end{aligned}$$

Since $\bar{B}_{n,d}-B_{n,d}=o_P(1)$ and $\zeta^*_{n,d}-\widehat{\zeta}_{n,d}=O_P(n^{-1/2})$ (Proposition B.1.2), the above is $O_P(n^{-1})$. Therefore,

$$\begin{aligned}n^{-1}\log\widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) &= n^{-1}\log\widetilde{L}_{n,d}(\zeta^*_{n,d})+\frac{1}{2}(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})^T\,\bar{B}_{n,d}\,(\zeta^*_{n,d}-\widehat{\zeta}_{n,d})\\ &= n^{-1}\log\widetilde{L}_{n,d}(\zeta^*_{n,d})+Q_n\end{aligned}$$

with $Q_n=O_P(n^{-1})$ or equivalently

$$\log\widetilde{L}_{n,d}(\widehat{\zeta}_{n,d})=\log\widetilde{L}_{n,d}(\zeta^*_{n,d})+Q^*_n$$

with $Q^*_n=O_P(1)$. Hence, for models M_{d_0} and M_d, we obtain ($d_0\neq d=1,\ldots,D$)

$$\begin{aligned}\Delta\mathsf{IC}_n(d_0,d) &= \mathsf{IC}_n(M_{d_0})-\mathsf{IC}_n(M_d)\\ &= 2\{\log\widetilde{L}_{n,d}(\widehat{\zeta}_{n,d})-\log\widetilde{L}_{n,d_0}(\widehat{\zeta}_{n,d_0})\}-\{\mathsf{pen}(n,d)-\mathsf{pen}(n,d_0)\}\\ &= 2\{\log\widetilde{L}_{n,d}(\zeta^*_{n,d})-\log\widetilde{L}_{n,d_0}(\zeta^*_{n,d_0})\}+Q^*_n\\ &\quad -\{\mathsf{pen}(n,d)-\mathsf{pen}(n,d_0)\}\\ &= R_n-P_n\end{aligned}$$

where $R_n=2\{\log\widetilde{L}_{n,d}(\zeta^*_{n,d})-\log\widetilde{L}_{n,d_0}(\zeta^*_{n,d_0})\}+Q^*_n$ and $P_n=\mathsf{pen}(n,d)-\mathsf{pen}(n,d_0)$.

The former holds for all models M_{d_0} and M_d ($d_0\neq d=1,\ldots,D$), thus also for all models $M_{d_0}\in\widetilde{\mathcal{J}}_0$ and $M_d\in\widetilde{\mathcal{J}}\backslash\widetilde{\mathcal{J}}_0$. We then have

$$\max_{M_d\notin\widetilde{\mathcal{J}}_0}\Delta\mathsf{IC}_n(d_0,d)=R^*_n-P^*_n$$

where R^*_n and P^*_n are R_n resp. P_n corresponding to $M_{d_0}\in\widetilde{\mathcal{J}}_0$ and $M_d\in\widetilde{\mathcal{J}}\backslash\widetilde{\mathcal{J}}_0$ for which $\Delta\mathsf{IC}_n(d_0,d)$ is maximal.

Further, (5.11) implies that for all $\xi>0$ there exists $\Delta(\xi)<\infty$ and $N_1(\xi)$ such that for all $n>N_1(\xi)$: $P(R^*_n>\Delta(\xi))<\xi/2$. Also, due to the penalty condition, there exists $N_2(\xi)$ such that for all $n>N_2(\xi)$: $P(P^*_n<\Delta(\xi))<\xi/2$.

Let $n > \max(N_1(\xi), N_2(\xi))$, then

$$\begin{aligned} P(\max_{M_d \notin \widetilde{\mathcal{J}}_0} \Delta\mathsf{IC}_n(d_0, d) < 0) &= P(R_n^* < P_n^*) \\ &\geq P(R_n^* \leq \Delta(\xi), P_n^* \geq \Delta(\xi)) \\ &\geq 1 - P(R_n^* < \Delta(\xi)) - P(P_n^* < \Delta(\xi)) \\ &> 1 - \xi. \end{aligned}$$

Consequently, $\lim_{n\to\infty} P(\max_{M_d \notin \widetilde{\mathcal{J}}_0} \Delta\mathsf{IC}_n(d_0, d) < 0) = 1$. □

The penalty in AIC does not depend on the sample size, hence Proposition 5.3.2 implies that model selection based on AIC is not consistent. The penalty in BIC includes the sample size and by Proposition 5.3.2 model selection based on BIC is consistent. A possible drawback of using a sample size dependent penalty is that, due to an increase in penalty, simpler models tend to be selected for a larger sample size, which might be counter-intuitive.

5.4 Illustrative example

We apply the discussed methodology to the udder infection data described in Section 1.2.1.

Consider the udder infection data - version 2 (407 cows). We construct 4 EAC's, 4 FNAC's and 4 PNAC's (Table 5.3). Figure 5.1 visualizes the induced dependence structures. We also build 64 Joe-Hu copulas (Table 5.4). The considered association patterns are depicted in Figure 5.2. For the Joe-Hu copulas, we have to specify for which pairs of udder quarters extra dependence is added on top of the overall association captured by the Laplace transform ψ, i.e., we need to identify the pairs (i, j) for which K_{ij} is different from the independence copula. The situation is summarized in Table 5.2. Further, we take $\nu_i = 0$ $(i = 1, \ldots, 4)$ and, within a specific dependence pattern, all K_{ij} not equal to the independence copula are of the same copula type (e.g., Clayton). By doing so, a nested sequence of Joe-Hu models is obtained. It is possible to combine K_{ij}'s of different copula types (e.g., Clayton and Gumbel), but since this would lead to an increased number of possible copulas, we do not consider this option. Note that the dependence pattern in Figure 5.2 (b) is the same as the one of PNAC in Figure 5.1 (c). However, the copula underlying the pattern in Figure 5.2 (b) allows a more flexible mixing of copula components (Section 5.1).

The obtained AIC and BIC values are listed in Table 5.3 and Table 5.4, with the three best models marked in bold. It follows that for the udder infection data the simpler

Table 5.2: Considered Joe-Hu structures for the udder infection data.

pattern	pairs (i,j) for which $K_{ij}(u,v) \neq uv$	number of parameters
a	(1,2)	2
b	(1,2), (3,4) with $K_{12} \neq K_{34}$	3
c	(1,2), (3,4), (1,3), (2,4) with $K_{12} \neq K_{34} \neq K_{13} = K_{24}$	4
d	(1,2), (3,4), (1,3), (2,4), (2,3) with $K_{12} \neq K_{34} \neq K_{13} = K_{24} \neq K_{23}$	5

Table 5.3: AIC/BIC-values for EAC, FNAC and PNAC.

ψ	EAC	FNAC	PNAC
Clayton	307.38 / 311.38	308.72 / 320.75	305.85 / 317.88
Gumbel	386.75 / 390.76	372.05 / 384.07	358.40 / 370.43
Frank	315.23 / 319.23	312.66 / 324.69	307.72 / 319.74
Joe	440.61 / 444.62	421.09 / 433.11	405.67 / 417.70

models, i.e., EAC, FNAC and PNAC are insufficient. It is the Joe-Hu copula with the most elaborate dependence structure, pattern d, that outperforms all other models. The parameter estimates of the preferred model are listed in the left panel of Table 5.5, while the corresponding estimated lower and upper tail dependence values are displayed in the right panel. The results in Section 5.7 are applied. The tail dependence coefficients for udder pair $(i,j) \neq (1,4)$ are calculated using the estimated parameter of the Laplace transform (θ^C) and the estimated value of the parameter θ^J_{ij} of K_{ij} ($i < j \in \{1,2,3,4\}$). Since K_{14} equals the independence copula, the tail dependence coefficients for udder pair $(1,4)$ are calculated using only the estimated parameter of the Laplace transform (θ^C). To obtain standard errors 1000 bootstrap samples are used; the resampling algorithms (parametric bootstrap) are given in Appendix A.2. From Table 5.5 it follows that the infection times are substantially correlated, i.e., the lower tail dependence equals 0.74 for all udder quarter pairs and the upper tail dependence of the udder quarter pairs ranges from 0 to 0.31. Late event times thus follow a similar association pattern, while for early event times the association changes per udder quarter pair. No specific symmetries are present.

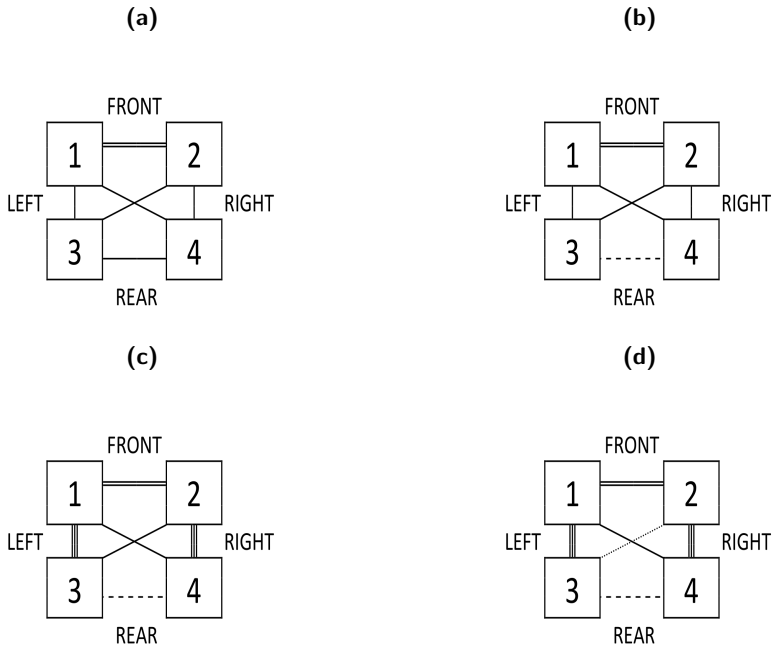


Figure 5.2: Considered Joe-Hu structures for the udder infection data - different line types represent different dependencies.

Table 5.4: AIC/BIC-values for Joe-Hu copulas.

ψ	K_{ij}	pattern a	pattern b	pattern c	pattern d
Clayton	Clayton	283.46 / 291.48	252.54 / 264.56	232.89 / 248.92	228.80 / 248.84
	Gumbel	279.44 / 287.46	254.75 / 266.78	214.56 / 230.59	**199.13** / **219.17**
	Frank	267.85 / 275.87	217.24 / 229.26	202.44 / 218.48	**198.16** / **218.21**
	Joe	280.43 / 288.44	256.34 / 268.36	214.02 / 230.06	**197.99** / **218.04**
Gumbel	Clayton	352.45 / 360.47	311.70 / 323.72	272.61 / 288.64	262.12 / 282.16
	Gumbel	355.27 / 363.29	326.87 / 338.89	271.07 / 287.10	247.56 / 267.60
	Frank	354.63 / 362.65	327.53 / 339.56	289.13 / 305.17	267.22 / 287.26
	Joe	355.85 / 363.86	327.96 / 339.99	271.86 / 287.89	248.54 / 268.58
Frank	Clayton	291.69 / 299.71	261.45 / 273.48	238.21 / 254.25	232.33 / 252.37
	Gumbel	287.43 / 295.45	262.29 / 274.32	221.71 / 237.74	205.12 / 225.17
	Frank	289.48 / 297.50	281.57 / 293.60	272.15 / 288.18	273.58 / 293.62
	Joe	287.48 / 295.50	262.36 / 274.38	221.68 / 237.72	205.03 / 225.07
Joe	Clayton	397.31 / 405.33	349.35 / 361.38	298.41 / 314.45	284.41 / 304.46
	Gumbel	402.36 / 410.37	369.28 / 381.31	301.67 / 317.71	275.06 / 295.10
	Frank	398.69 / 406.71	378.79 / 390.81	329.38 / 345.42	293.10 / 313.14
	Joe	402.89 / 410.91	370.26 / 382.28	302.59 / 318.62	276.21 / 296.26

Table 5.5: Estimates of the Clayton Laplace parameter (θ^C) and of the bivariate Joe copula parameters (θ^J_{ij}) as well as the corresponding estimated lower ($\widehat{\delta}_L$) and upper tail dependence parameter ($\widehat{\delta}_U$); the estimated standard errors se_1 and se_2 are obtained using the bootstrap Algorithms 1 and 2 (Appendix A.2).

	estimated parameter (se_1, se_2)	pair of udder quarters	$\widehat{\delta}_L$ (se_1, se_2)	$\widehat{\delta}_U$ (se_1, se_2)
θ^C	2.31 (0.30, 0.29)	1-4	0.74 (0.03, 0.03)	
θ^J_{12}	2.62 (0.73, 0.69)	1-2	0.74 (0.03, 0.03)	0.23 (0.03, 0.03)
θ^J_{23}	1.38 (0.23, 0.23)	2-3	0.74 (0.03, 0.03)	0.12 (0.04, 0.04)
θ^J_{34}	4.44 (1.03, 1.05)	3-4	0.74 (0.03, 0.03)	0.28 (0.02, 0.02)
$\theta^J_{13} = \theta^J_{24}$	9.65 (1.35, 1.37)	1-3 and 2-4	0.74 (0.03, 0.03)	0.31 (0.01, 0.01)

5.5 Simulation study

To allow flexible choices for ψ and K_{ij} $(1 \leq i < j \leq 4)$ in a Joe-Hu copula, we developed a generic R-program based on the copula formula in (5.5). Due to this generic character, some numerical care is needed, e.g., evaluation of the exponent of the negative inverse Laplace transform might be tedious due to limited precision. We addressed this issue by high precision calculations, i.e., instead of using the double precision numbers in R, we use multiple precision floating point numbers (Maechler, 2014). Further, note that the likelihood expression in (5.7) contains (higher order) partial derivatives of the copula. The exact expressions are typically quite cumbersome to obtain. In the program, we therefore used finite forward differences as approximations. To evaluate the numerical performance of the generic R-program, we set up a small simulation study.

5.5.1 Simulation setting

We generate 600 datasets, each containing 500 clusters of size 4 from either the Clayton copula with $\theta = 3.19$, the Frank copula with $\theta = 5.76$ or the Joe-Hu copula

$$\begin{aligned} & C(u_1, \ldots, u_4) \\ = \; & \left(1 + \left\{\left(u_1^{-\theta} - 1\right)^{\alpha} + \left(u_2^{-\theta} - 1\right)^{\alpha}\right\}^{\frac{1}{\alpha}} + \left\{\left(u_3^{-\theta} - 1\right)^{\alpha} + \left(u_4^{-\theta} - 1\right)^{\alpha}\right\}^{\frac{1}{\alpha}}\right)^{-\frac{1}{\theta}} \end{aligned}$$

with $\theta = 2.91$ and $\alpha = 1.17$. This Joe-Hu copula can be constructed by taking $\nu_1 = \ldots = \nu_4 = -2$, ψ a Clayton Laplace transform with parameter θ, $K_{12} = K_{34}$ a bivariate Gumbel copula with parameter α and all other K_{ij} independence copulas $(1 \leq i < j \leq 4)$. Based on the results in Section 5.7, it holds that the bivariate margins of the Clayton copula have lower tail correlation with $\delta_L = 0.81$, while those of the Frank copula display no tail dependence. The $(1, 2)$ and $(3, 4)$ margins of the Joe-Hu copula have lower and upper tail correlation given by $\delta_{L_{12}} = \delta_{L_{34}} = 0.82$, resp. $\delta_{U_{12}} = \delta_{U_{34}} = 0.19$, all other bivariate Joe-Hu margins show only lower tail dependence with $\delta_L = 0.79$.

To obtain event times T_{sr} $(s = 1, \ldots, 500$ and $r = 1, \ldots, 4)$ Weibull margins with scale $\lambda = 0.5$ and shape $\rho = 1.5$ are used. The censoring mechanism is assumed to be univariate, i.e., $C_{sr} = C_s$ $(s = 1, \ldots, 500$ and $r = 1, \ldots, 4)$, Weibull with scale and shape given by $\lambda = 0.15, \rho = 1.5$, resp. $\lambda = 0.85, \rho = 1.5$, leading to approximately 23%, resp. 63% censoring. The latter corresponds to the censoring present in the udder infection data. The observed data can then be calculated as $Y_{sr} = \min(T_{sr}, C_{sr})$ and $\delta_{sr} = I(T_{sr} \leq C_{sr})$ $(s = 1, \ldots, 500$ and $r = 1, \ldots, 4)$.

Table 5.6: Simulation results. The first and second column give the simulation setting. The third column reveals the used stepsize. Columns four and five list the mean estimated copula parameters, while columns six to eight contain the mean estimated lower ($\widehat{\delta}_L$, $\widehat{\delta}_{U_{12}} = \widehat{\delta}_{L_{34}}$) and upper ($\widehat{\delta}_{U_{12}} = \widehat{\delta}_{L_{34}}$) tail dependence parameter. All estimated parameters are supplemented with the corresponding empirical standard deviation.

copula	censoring	stepsize	$\widehat{\theta}$ (se)	$\widehat{\alpha}$ (se)	$\widehat{\delta}_{L_{12}} = \widehat{\delta}_{L_{34}}$ (se)	$\widehat{\delta}_L$ (se)	$\widehat{\delta}_{U_{12}} = \widehat{\delta}_{U_{34}}$ (se)
Clayton	23%	0.001	3.09 (0.23)			0.80 (0.01)	
		0.0005	3.09 (0.23)			0.80 (0.01)	
	63%	0.001	3.04 (0.33)			0.79 (0.02)	
		0.0005	3.04 (0.33)			0.79 (0.02)	
Frank	23%	0.001	5.74 (0.27)				
		0.0005	5.73 (0.26)				
	63%	0.001	5.72 (0.38)				
		0.0005	5.72 (0.37)				
Joe-Hu	23%	0.001	2.81 (0.23)	1.18 (0.03)	0.81 (0.01)	0.78 (0.02)	0.20 (0.03)
		0.0005	2.81 (0.23)	1.18 (0.03)	0.81 (0.01)	0.78 (0.02)	0.20 (0.03)
	63%	0.001	2.77 (0.33)	1.18 (0.04)	0.81 (0.02)	0.78 (0.02)	0.20 (0.04)
		0.0005	2.78 (0.32)	1.18 (0.04)	0.81 (0.02)	0.78 (0.02)	0.20 (0.04)

5.5.2 Simulation results

The results, obtained by applying the generic R-program with a stepsize of either 0.001 or 0.0005 for the finite forward differences, are summarized in Table 5.6. It follows that, on average and taking the empirical standard deviation into account, the estimation of the copula and tail dependence parameters is on target. The performance is somewhat more accurate for the Frank copula and slightly better for light censored data.

5.6 Discussion

In this chapter, we show that Joe-Hu copulas allow a flexible modeling of diverse dependence structures, including the more restricted ones implied by the exchangeable and the nested Archimedean copulas. Vine copulas also constitute a flexible alternative (Aas *et al.*, 2009; Berg and Aas, 2009; Kurowicka and Joe, 2011). The construction of the latter is based on a decomposition of the joint copula density into a cascade of bivariate (un)conditional copula densities. Given the variety of possible bivariate copula

densities and the numerous ways of decomposing the joint copula density, it is clear that vine copulas can describe multiple (non-symmetric) association patterns. Compared to a Joe-Hu copula, a vine copula has no closed form. Hence, the calculation of bivariate margins and tail dependence coefficients is non-trivial. Further, even though vine copulas are well explored for complete data, no extension to right-censored data is yet available. This is the subject of future research.

Further, we prove that, under certain conditions, model selection based on information criterion (5.8) is weakly consistent or consistent. Sin and White (1996) state conditions on the penalty under which a selection criterion of the form (5.8) is strongly consistent, i.e., under which the best (parsimonious) model is selected almost surely. The necessary tool to establish the latter is the law of the iterated logarithm. A translation to the context of (misspecified) semiparametric copulas might be possible, but is beyond the scope of this dissertation. An alternative selection method for semiparametric copulas is the copula information criterion (CIC) by Grønneberg and Hjort (2014). They construct an unbiased estimator of the Kullback-Leibler discrepancy between a model and the true copula C, leading to some extra penalty terms as compared to (5.8). Since the derivation has only been carried out for complete data, CIC is not applicable in the current setting.

5.7 Addendum

In this section we explore the tail behavior of the copulas discussed in Section 5.1.

Consider a four-variate exchangeable or nested Archimedean copula (5.1) - (5.3). The bivariate copula margins are given by

$$C(u_i, u_j) = \widetilde{\psi}(\widetilde{\psi}^{-1}(u_i) + \widetilde{\psi}^{-1}(u_j)) \tag{5.12}$$

where $\widetilde{\psi} \in \{\psi, \psi_1, \psi_2, \psi_3\}$, i.e., each bivariate margin is an exchangeable Archimedean copula. The tail dependence parameters of (5.12) can be calculated via Theorem B.1.1. For $\widetilde{\psi} \in \{$Clayton, Gumbel, Frank, Joe$\}$ the result is contained in Table 5.7. Figure 5.3 visualizes the induced association for $\widetilde{\psi} \in \{$Clayton, Gumbel, Frank, Joe$\}$ with parameter $\theta = 3$. It follows that a Clayton copula is lower tail dependent, a Gumbel and a Joe copula have upper tail dependence, while a Frank copula exhibits no tail dependence.

Further, consider two four-variate Joe-Hu copulas (5.5): one with $\nu_i = 0$ or $p_i = 1/3$ for all $i \in \{1, \dots, 4\}$ (as in Section 5.4) and one with $\nu_i = -2$ or $p_i = 1$ for all $i \in \{1, \dots, 4\}$ (as in Section 5.5). From (5.6) we get that the first Joe-Hu copula has bivariate margins

given by

$$C(u_i,u_j) = \psi\big(-\log K_{ij}(\exp(-1/3\psi^{-1}(u_i)),\exp(-1/3\psi^{-1}(u_j))) + 2/3\psi^{-1}(u_i)+2/3\psi^{-1}(u_j)\big) \tag{5.13}$$

while those of the second Joe-Hu copula take the form

$$C(u_i,u_j) = \psi\big(-\log K_{ij}(\exp(-\psi^{-1}(u_i)),\exp(-\psi^{-1}(u_j)))\big). \tag{5.14}$$

The copulas in (5.13) and (5.14) are used to demonstrate the gain in association flexibility compared to a bivariate exchangeable Archimedean copula as well as to illustrate the impact of the chosen ν_i on the modeled association.

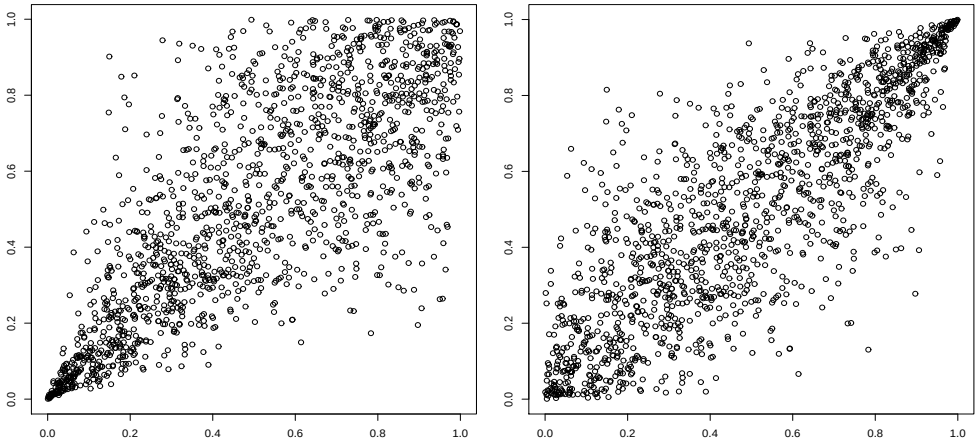
For various choices of ψ and K_{ij}, the tail dependence parameters of (5.13) and (5.14) can be calculated via Theorem B.1.2 and Theorem B.1.3, respectively. The results are summarized in Table 5.8 and Table 5.9. Figure 5.4 and Figure 5.5 allow a visual comparison of the induced association for ψ the Clayton Laplace transform with $\theta = 3$ and K_{ij} a max-id copula with parameter $\alpha = 4$ where $K_{ij} \in \{$Clayton, Gumbel, Frank, Joe$\}$.

It appears that the lower tail dependence caused by ψ a Clayton Laplace transform is strengthened for K_{ij} a Clayton copula and that some upper tail dependence is introduced by taking K_{ij} either a Gumbel or a Joe copula. Setting K_{ij} to be a Frank copula has no effect on the modeled tail dependence. Moreover, if all other model parameters are taken constant, the chosen value of ν_i seems to impact the strength of the modeled correlation, i.e., a smaller value of ν_i leads to a somewhat stronger association. In practice ν_i is fixed and consequently the parameters of ψ and K_{ij} will have to accommodate for the choice of ν_i.

Note that the above results are to be expected: (5.12) is an one-parameter copula, while (5.13) and (5.14) are two-parameter models. Therefore, the former can describe either lower or upper tail dependence (if present), whereas the latter two can capture both lower and upper tail dependence (if present).

(a) $\delta_L = 0.79$ and $\delta_U = 0$

(b) $\delta_L = 0$ and $\delta_U = 0.74$



(c) $\delta_L = 0$ and $\delta_U = 0$

(d) $\delta_L = 0$ and $\delta_U = 0.74$

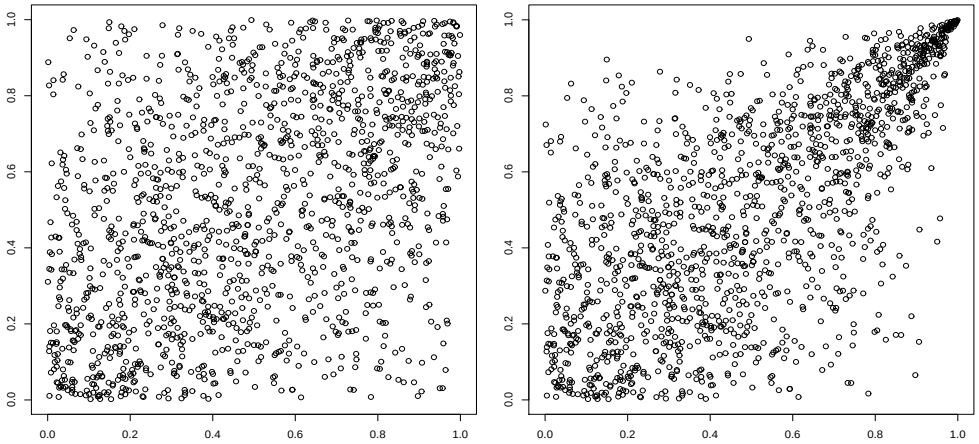
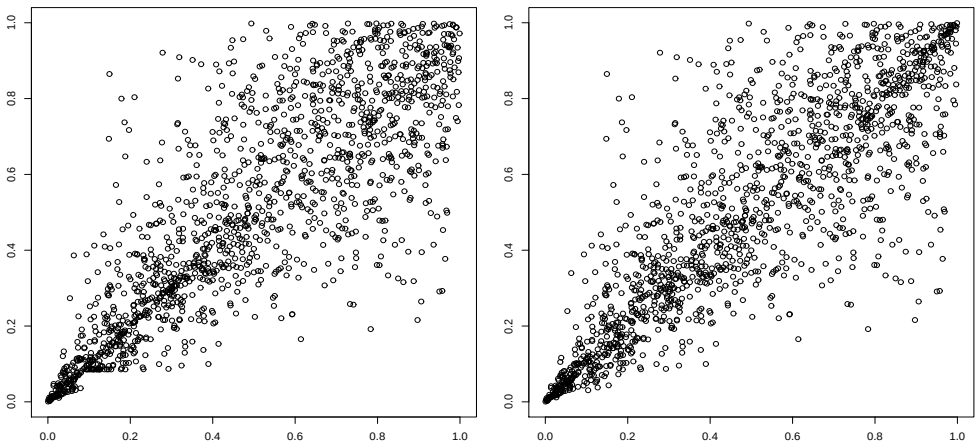


Figure 5.3: Tail dependencies for several bivariate exchangeable Archimedean copulas defined by (5.12) with $\theta = 3$: (a) Clayton, (b) Gumbel, (c) Frank, (d) Joe.

(a) $\delta_L = 0.84$ and $\delta_U = 0$

(b) $\delta_L = 0.83$ and $\delta_U = 0.27$



(c) $\delta_L = 0.79$ and $\delta_U = 0$

(d) $\delta_L = 0.79$ and $\delta_U = 0.27$

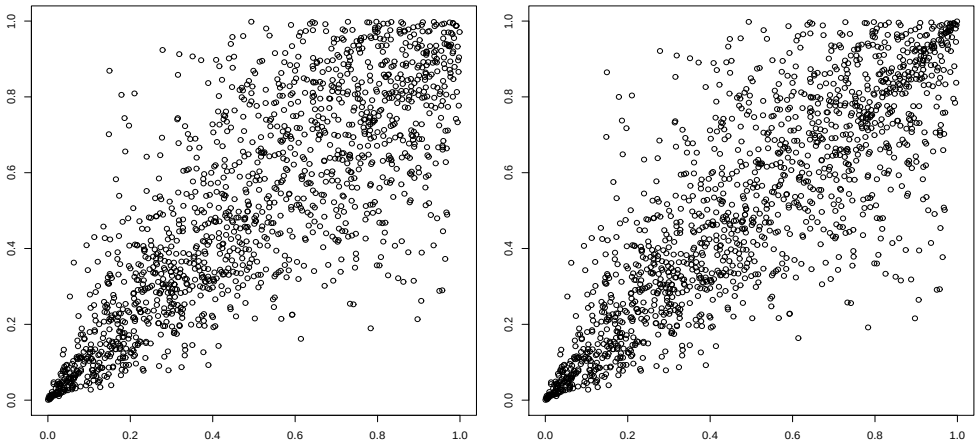
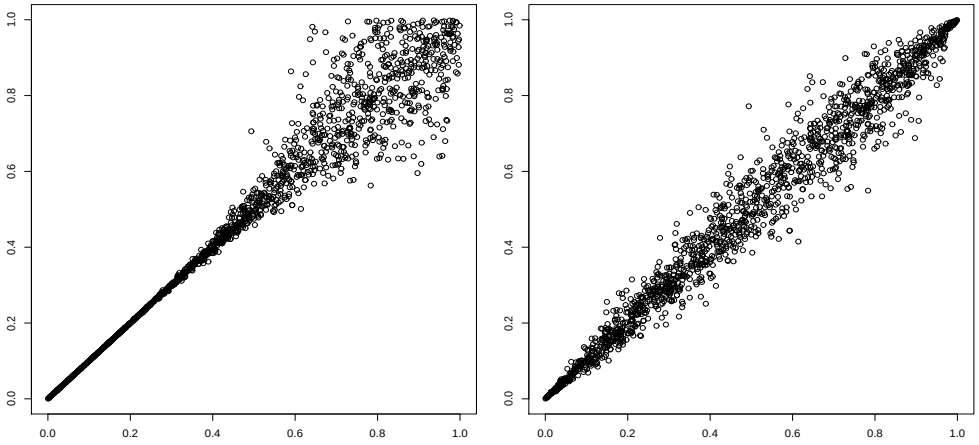


Figure 5.4: Tail dependencies for several bivariate Joe-Hu copulas defined by (5.13) with $\theta = 3$ and $\alpha = 4$, (a) K_{ij} = Clayton, (b) K_{ij} = Gumbel, (c) K_{ij} = Frank, (d) K_{ij} = Joe.

(a) $\delta_L = 1$ and $\delta_U = 0$

(b) $\delta_L = 0.94$ and $\delta_U = 0.81$



(c) $\delta_L = 0.79$ and $\delta_U = 0$

(d) $\delta_L = 0.79$ and $\delta_U = 0.81$

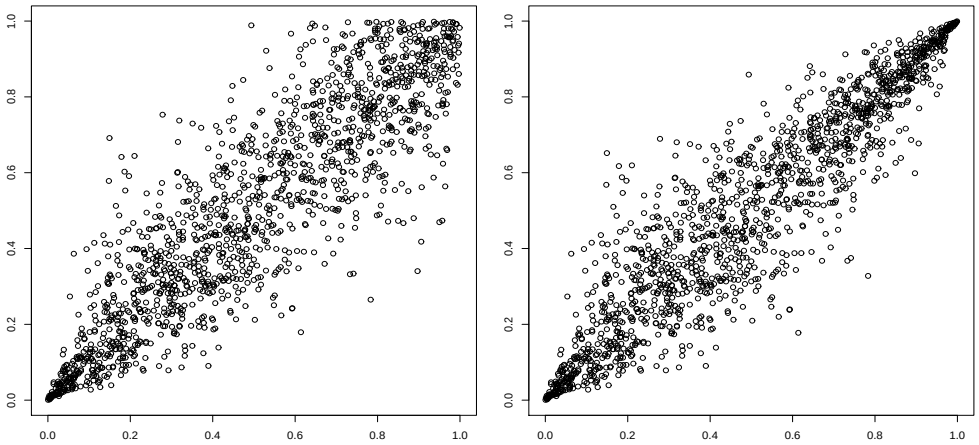


Figure 5.5: Tail dependencies for several bivariate Joe-Hu copulas defined by (5.14) with $\theta = 3$ and $\alpha = 4$, (a) K_{ij} = Clayton, (b) K_{ij} = Gumbel, (c) K_{ij} = Frank, (d) K_{ij} = Joe.

Table 5.7: Lower (δ_L) and upper (δ_U) tail dependence parameter for some bivariate exchangeable Archimedean copulas defined by (5.12).

copula	δ_L	δ_U
Clayton	$2^{-1/\theta}$	0
Gumbel	0	$2-2^{1/\theta}$
Frank	0	0
Joe	0	$2-2^{1/\theta}$

Table 5.8: Lower (δ_L) and upper (δ_U) tail dependence parameter for some bivariate Joe-Hu copulas defined by (5.13).

ψ	K	δ_L	δ_U
Clayton	Clayton	$\left(\frac{5}{3}\right)^{-1/\theta}$	0
	Gumbel	$\left(\frac{1}{3}(4+2^{1/\alpha})\right)^{-1/\theta}$	$\frac{1}{3}(2-2^{1/\alpha})$
	Frank	$2^{-1/\theta}$	0
	Joe	$2^{-1/\theta}$	$\frac{1}{3}(2-2^{1/\alpha})$
Gumbel	Clayton	0	$2-2^{1/\theta}$
	Gumbel	0	$2-\left(\frac{1}{3}(4-2^{1/\alpha})\right)^{1/\theta}$
	Frank	0	$2-2^{1/\theta}$
	Joe	0	$2-\left(\frac{1}{3}(4-2^{1/\alpha})\right)^{1/\theta}$
Frank	Clayton	0	0
	Gumbel	0	$\frac{1}{3}(2-2^{1/\alpha})$
	Frank	0	0
	Joe	0	$\frac{1}{3}(2-2^{1/\alpha})$
Joe	Clayton	0	$2-2^{1/\theta}$
	Gumbel	0	$2-\left(\frac{1}{3}(4-2^{1/\alpha})\right)^{1/\theta}$
	Frank	0	$2-2^{1/\theta}$
	Joe	0	$2-\left(\frac{1}{3}(4-2^{1/\alpha})\right)^{1/\theta}$

Table 5.9: Lower (δ_L) and upper (δ_U) tail dependence parameter for some bivariate Joe-Hu copulas defined by (5.14).

ψ	K	δ_L	δ_U
Clayton	Clayton	1	0
	Gumbel	$2^{-1/\theta\alpha}$	$2-2^{1/\alpha}$
	Frank	$2^{-1/\theta}$	0
	Joe	$2^{-1/\theta}$	$2-2^{1/\alpha}$
Gumbel	Clayton	$2^{-1/\alpha}$ if $\theta=1$ and 1 if $\theta>1$	$2-2^{1/\theta}$
	Gumbel	0	$2-2^{1/\theta\alpha}$
	Frank	0	$2-2^{1/\theta}$
	Joe	0	$2-2^{1/\theta\alpha}$
Frank	Clayton	$2^{-1/\alpha}$	0
	Gumbel	0	$2-2^{1/\alpha}$
	Frank	0	0
	Joe	0	$2-2^{1/\alpha}$
Joe	Clayton	$2^{-1/\alpha}$	$2-2^{1/\theta}$
	Gumbel	0	$2-2^{1/\theta\alpha}$
	Frank	0	$2-2^{1/\theta}$
	Joe	0	$2-2^{1/\theta\alpha}$

Chapter 6

Nonparametric copula estimation

In Chapter 5 semiparametric copulas are used to describe clustered right-censored event time data. In this chapter, the focus is on bivariate nonparametric copula estimation, i.e., the marginal survival functions and the copula function are modeled in a nonparametric way.

For complete time-to-event data, a nonparametric copula estimator has been introduced and studied by Deheuvels (1979) as well as Gaenssler and Stute (1987). More recent papers include Fermanian *et al.* (2004) and Segers (2012). For right-censored event time data, less work has been done. Recently, Gribkova and Lopez (2014) defined a nonparametric copula estimator for various right-censoring schemes. In this chapter, an alternative is proposed. We prove consistency and derive an asymptotic i.i.d. representation. The ideas in Lin and Ying (1993) as well as Wang and Wells (1997) are instrumental.

In Section 6.1 we give the right-censoring schemes that are considered. Section 6.2 contains existing nonparametric survival function estimators. In Section 6.3 a new nonparametric copula estimator is defined. Consistency is established in Section 6.4, an asymptotic i.i.d. representation is derived in Section 6.5. In Section 6.6 we use a simulation study to investigate the finite sample performance of the new estimator. A comparison with the nonparametric copula estimator of Gribkova and Lopez (2014) is provided. Section 6.7 closes the chapter with a discussion. Section 6.8 contains supplementary material. Key results needed for the proofs are collected in Appendix B.2.

6.1 Censoring schemes

Consider bivariate right-censored event time data. Denote the observed time of item r $(r = 1, 2)$ in cluster s $(s = 1, \ldots, n)$ by $Y_{sr} = \min(T_{sr}, C_{sr})$ with T_{sr} the true event time and C_{sr} the censoring time. The indicator $\delta_{sr} = I(T_{sr} \leq C_{sr})$ equals one if $Y_{sr} = T_{sr}$ and zero otherwise. Event times and censoring times are assumed to be independent.

The vector (T_1, T_2) of event times has joint survival function S, the marginal survival and distribution functions are denoted by S_1 and S_2, resp. F_1 and F_2. The vector (C_1, C_2) of censoring times has joint survival function S_C, the marginal survival and distribution functions are given by S_{G_1} and S_{G_2}, resp. G_1 and G_2. The joint survival function of (Y_1, Y_2) is denoted by S_Y, where $Y_r = \min(T_r, C_r)$ $(r = 1, 2)$.

Sklar's theorem implies that $S(t_1, t_2) = C(S_1(t_1), S_2(t_2))$ for a copula function C. We aim at estimating C in a nonparametric way. Two settings for (C_1, C_2) are considered:

(1) univariate censoring, i.e., $C_1 = C_2 = \bar{C}$ and thus

$$S_C(c_1, c_2) = P(C_1 > c_1, C_2 > c_2) = S_G(\max(c_1, c_2))$$

where S_G is the survival function of $\bar{C}$. In practice, S_G is replaced by the Kaplan-Meier estimator $\widehat{S}_G$ based on $(\widetilde{Y}_s, 1 - \widetilde{\delta}_s)$ where $\widetilde{Y}_s = \max(Y_{s1}, Y_{s2})$ and $\widetilde{\delta}_s = \delta_{s1}\delta_{s2}$ $(s = 1, \ldots, n)$. Indeed, $\bar{C}$ can be seen as the censoring time of $\widetilde{T} = \max(T_1, T_2)$. We then observe $\widetilde{Y}_s = \min(\widetilde{T}_s, \bar{C}_s) = \min(\max(T_{s1}, T_{s2}),\ \bar{C}_s) = \max(\min(T_{s1}, \bar{C}_s), \min(T_{s2}, \bar{C}_s)) = \max(Y_{s1}, Y_{s2})$ and $\widetilde{\delta}_s = I(\widetilde{T}_s \leq \bar{C}_s) = I(\max(T_{s1}, T_{s2}) \leq \bar{C}_s) = I(T_{s1} \leq \bar{C}_s)\, I(T_{s2} \leq \bar{C}_s) = \delta_{s1}\delta_{s2}$ $(s = 1, \ldots, n)$.

(2) copula censoring, i.e., C_1 and C_2 are associated via a copula $\widetilde{C}$ and thus

$$S_C(c_1, c_2) = P(C_1 > c_1, C_2 > c_2) = \widetilde{C}(S_{G_1}(c_1), S_{G_2}(c_2)).$$

In practice, S_{G_r} is replaced by the Kaplan-Meier estimator $\widehat{S}_{G_r}$ based on $(Y_{sr}, 1 - \delta_{sr})$ $(s = 1, \ldots, n$ and $r = 1, 2)$ and $\widetilde{C}$ is assumed to be known, i.e., $\widetilde{C} = \widetilde{C}_\theta$ with known parameter θ. We also discuss the case of an unknown parameter θ. The setting of C_1 and C_2 independent is covered by taking $\widetilde{C}$ the independence copula.

6.2 Nonparametric survival function estimators

In this section two nonparametric survival function estimators for right-censored event time data are given.

Lopez and Saint-Pierre (2012) as well as Gribkova *et al.* (2013) consider an estimator of the form

$$S_n(t_1, t_2) = \frac{1}{n} \sum_{s=1}^{n} W_{sn} I(Y_{s1} > t_1, Y_{s2} > t_2) \tag{6.1}$$

where $W_{sn} = \delta_{s1}\delta_{s2}/\widehat{S}_C^{lc}(Y_{s1}, Y_{s2})$ and $\widehat{S}_C^{lc}$ is an appropriate estimator for a left-continuous version of S_C. W_{sn} is non-zero if cluster s contains only events and zero otherwise ($s = 1, \ldots, n$).

Alternatively, Lin and Ying (1993) as well as Wang and Wells (1997) consider

$$\begin{aligned} S_Y(t_1, t_2) &= P(Y_1 > t_1, Y_2 > t_2) \\ &= P(T_1 > t_1, T_2 > t_2)\, P(C_1 > t_1, C_2 > t_2) \\ &= S(t_1, t_2)\, S_C(t_1, t_2) \end{aligned}$$

or

$$S(t_1, t_2) = \frac{S_Y(t_1, t_2)}{S_C(t_1, t_2)}$$

to obtain

$$\widehat{S}(t_1, t_2) = \frac{n^{-1} \sum_{s=1}^{n} I(Y_{s1} > t_1, Y_{s2} > t_2)}{\widehat{S}_C(t_1, t_2)} \tag{6.2}$$

with $\widehat{S}_C$ an appropriate estimator for S_C. In contrast to the estimator $S_n(t_1, t_2)$, all clusters can contribute to the estimator $\widehat{S}(t_1, t_2)$.

For complete data both estimators coincide with the empirical survival function.

6.3 Nonparametric copula function estimators

In this section two nonparametric copula estimators for right-censored event time data are defined.

Given the estimator in (6.1), Gribkova and Lopez (2014) apply the Sklar's theorem to obtain

$$C_n(u_1, u_2) = S_n(S_{n1}^{-1}(u_1), S_{n2}^{-1}(u_2)) \tag{6.3}$$

where $S_{n1}(t_1) = S_n(t_1, 0)$ and $S_{n2}(t_2) = S_n(0, t_2)$. The estimator contains the weights W_{sn} and thus ignores any cluster with at least one censored observation.

Alternatively, define $V_r = S_r(Y_r)$ $(r = 1, 2)$ and consider

$$\begin{aligned} & P(V_1 < u_1, V_2 < u_2) \\ &= P(T_1 > S_1^{-1}(u_1), T_2 > S_2^{-1}(u_2))\, P(C_1 > S_1^{-1}(u_1), C_2 > S_2^{-1}(u_2)) \\ &= C(u_1, u_2)\, S_C(S_1^{-1}(u_1), S_2^{-1}(u_2)) \end{aligned}$$

or

$$C(u_1, u_2) = \frac{P(V_1 < u_1, V_2 < u_2)}{S_C(S_1^{-1}(u_1), S_2^{-1}(u_2))} \tag{6.4}$$

to obtain

$$\widehat{C}(u_1, u_2) = \frac{n^{-1} \sum_{s=1}^{n} I(\widehat{V}_{s1} < u_1, \widehat{V}_{s2} < u_2)}{\widehat{S}_C(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2))} \tag{6.5}$$

where $\widehat{V}_{sr} = \widehat{S}_r(Y_{sr})$, $\widehat{S}_r$ is the Kaplan-Meier estimator for S_r based on (Y_{sr}, δ_{sr}) and $\widehat{S}_r^{-1}$ is the corresponding quantile function ($s = 1, \ldots, n$ and $r = 1, 2$). Further, $\widehat{S}_C$ is an appropriate estimator for S_C. Note that, unlike the estimator $C_n(u_1, u_2)$, all clusters can contribute to the estimator $\widehat{C}(u_1, u_2)$.

For complete data both estimators reduce to the empirical copula function.

Before we explore the asymptotic behavior of the nonparametric copula estimator (6.5), some extra notation is introduced. Denote the upper endpoint of the support of any distribution function L by T_L, i.e., $T_L = \inf\{t : L(t) = 1\}$. Due to independence of T_r and C_r, it holds that $T_{H_r} = \min(T_{F_r}, T_{G_r})$ for the distribution functions F_r of T_r, G_r of C_r and H_r of Y_r $(r = 1, 2)$. Further, in the case of univariate censoring, $T_{\widetilde{H}} = \min(T_{\widetilde{F}}, T_G)$ where $\widetilde{F}, G$ and $\widetilde{H}$ are the distribution functions of resp. $\widetilde{T} = \max(T_1, T_2)$, $\bar{C}$ and $\widetilde{Y} = \min(\widetilde{T}, \bar{C})$.

Moreover, to study the asymptotic properties of the nonparametric copula estimator (6.5), we need the following conditions:

(C1) S and S_C are continuous.

(C2) S_Y is Lipschitz of order 1.

(C3) S_Y is differentiable and its partial derivatives $S_Y^{(1)}$ and $S_Y^{(2)}$ are Lipschitz of order 1.

(C4) S_{G_r} is Lipschitz of order 1 $(r = 1, 2)$.

(C5) S_G is Lipschitz of order 1.

(C6) S_G is differentiable and its derivative $S_G^{'}$ is Lipschitz of order 1.

(C7) $S_r^{-1}(u_r) < T_{H_r}$ $(r = 1, 2)$.

(C8) $S_r^{-1}(u_r) < T_{\widetilde{H}}$ $(r = 1, 2)$.

(C9) For $S_r(T_{G_r}) < u_r < 1$, S_r is differentiable at $S_r^{-1}(u_r)$ with $f_r(S_r^{-1}(u_r)) > 0$ where $f_r = -S_r^{'}$ $(r = 1, 2)$.

(C10) S_{G_r} is differentiable with $|S_{G_r}^{'}|$ bounded in $[0, T]$ with $T < T_{H_r}$ $(r = 1, 2)$.

(C11) S_G is differentiable with $|S_G^{'}|$ bounded in $[0, T]$ with $T < T_{\widetilde{H}}$.

6.4 Consistency

In (6.4) and (6.5), denote the numerator by A resp. $\widehat{A}$ and the denominator by B resp. $\widehat{B}$, i.e.,

$$C(u_1, u_2) = \frac{A}{B} \quad \text{and} \quad \widehat{C}(u_1, u_2) = \frac{\widehat{A}}{\widehat{B}}.$$

In this section the consistency of the estimator $\widehat{C}(u_1, u_2)$ is established by proving that $\widehat{A} \to A$ a.s. and that $\widehat{B} \to B$ a.s.

We start by showing that $\widehat{A} \to A$ a.s.

Theorem 6.4.1. *Assume (C1), (C2), (C7) and (C9). Then, as $n \to \infty$, $\widehat{A} \to A$ a.s.*

Proof. We have

$$\widehat{A} - A = A_{n1} + A_{n2} + A_{n3}$$

with

$$\begin{aligned}
A_{n1} &= \frac{1}{n}\sum_{s=1}^{n} I(V_{s1} < u_1, V_{s2} < u_2) - P(V_1 < u_1, V_2 < u_2) \\
A_{n2} &= \frac{1}{n}\sum_{s=1}^{n} I(\widehat{V}_{s1} < u_1, \widehat{V}_{s2} < u_2) - P(\widehat{V}_1 < u_1, \widehat{V}_2 < u_2) \\
&- \Big[\frac{1}{n}\sum_{s=1}^{n} I(V_{s1} < u_1, V_{s2} < u_2) - P(V_1 < u_1, V_2 < u_2)\Big] \\
A_{n3} &= P(\widehat{V}_1 < u_1, \widehat{V}_2 < u_2) - P(V_1 < u_1, V_2 < u_2).
\end{aligned}$$

By the strong law of large numbers: $A_{n1} \to 0$ a.s.

For A_{n2}, introduce the notation

$$S_{n,Y}(y_1, y_2) = \frac{1}{n} \sum_{s=1}^{n} I(Y_{s1} > y_1, Y_{s2} > y_2).$$

By continuity of S_1 and S_2, we can write

$$\frac{1}{n} \sum_{s=1}^{n} I(V_{s1} < u_1, V_{s2} < u_2) = S_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2)).$$

Using that the jumps of a Kaplan-Meier estimator are $O(n^{-1})$ a.s. (Aly *et al.*, 1985), we have

$$\frac{1}{n} \sum_{s=1}^{n} I(\widehat{V}_{s1} < u_1, \widehat{V}_{s2} < u_2) = S_{n,Y}(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2)) + O(n^{-1}) \text{ a.s.}$$

Further, note that

$$\begin{aligned} P(V_1 < u_1, V_2 < u_2) &= S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2)) \\ P(\widehat{V}_1 < u_1, \widehat{V}_2 < u_2) &= S_Y(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2)) + O(n^{-1}) \text{ a.s.} \end{aligned}$$

We therefore have

$$\begin{aligned} A_{n2} &= S_{n,Y}(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2)) - S_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2)) \\ &- \left[S_Y(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2)) - S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2)) \right] + O(n^{-1}) \text{ a.s.} \end{aligned}$$

Lemma B.2.2 and Lemma B.2.6 imply that $A_{n2} = O(n^{-3/4}(\log n)^{-3/4})$ a.s.

Further, Lemma B.2.2 and (C2) indicate that

$$P(Y_1 > \widehat{S}_1^{-1}(u_1), Y_2 > \widehat{S}_2^{-1}(u_2)) - P(Y_1 > S_1^{-1}(u_1), Y_2 > S_2^{-1}(u_2)) \to 0 \text{ a.s.}$$

Again using that the jumps of a Kaplan-Meier estimator are $O(n^{-1})$ a.s., it follows that $A_{n3} \to 0$ a.s. □

In Theorems 6.4.2 and 6.4.3 we show, for the censoring schemes described in Section 6.1, that $\widehat{B} \to B$ a.s.

Theorem 6.4.2 (univariate censoring)**.** *Assume (C1), (C5), (C8) and (C9). Then, as* $n \to \infty$, $\widehat{B} \to B$ *a.s.*

Proof. We have to show that

$$\widehat{S}_G(\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2))) \to S_G(\max(S_1^{-1}(u_1), S_2^{-1}(u_2))) \text{ a.s.}$$

Apply the equality $\max(a,b) = a + \frac{1}{2}[|a-b| - (a-b)]$ and Lemma B.2.2 to obtain

$$\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2)) \to \max(S_1^{-1}(u_1), S_2^{-1}(u_2)) \text{ a.s.} \tag{6.6}$$

Then use the decomposition

$$\begin{aligned} & \widehat{S}_G(\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2))) - S_G(\max(S_1^{-1}(u_1), S_2^{-1}(u_2))) \\ = \; & \widehat{S}_G(\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2))) - S_G(\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2))) \\ + \; & S_G(\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2))) - S_G(\max(S_1^{-1}(u_1), S_2^{-1}(u_2))) \\ = \; & B_{n1} + B_{n2}. \end{aligned}$$

Lemma B.2.1 and (C8) imply that $B_{n1} \to 0$ a.s., while (C5) and (6.6) lead to $B_{n2} \to 0$ a.s. □

Theorem 6.4.3 (copula censoring)**.** *Assume (C1), (C4), (C7), (C9) and (C10). Then, as $n \to \infty$, $\widehat{B} \to B$ a.s.*

Proof. We have to show that

$$\widetilde{C}(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) \to \widetilde{C}(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2))) \text{ a.s.}$$

By the Lipschitz property of copulas (Nelsen, 2006), it suffices to show that for $r = 1, 2$:

$$\widehat{S}_{G_r}(\widehat{S}_r^{-1}(u_r)) \to S_{G_r}(S_r^{-1}(u_r)) \text{ a.s.}$$

We have

$$\widehat{S}_{G_r}(\widehat{S}_r^{-1}(u_r)) - S_{G_r}(S_r^{-1}(u_r)) = I_{n1r} + I_{n2r} + I_{n3r}$$

with

$$\begin{aligned} I_{n1r} &= \widehat{S}_{G_r}(S_r^{-1}(u_r)) - S_{G_r}(S_r^{-1}(u_r)) \\ I_{n2r} &= S_{G_r}(\widehat{S}_r^{-1}(u_r)) - S_{G_r}(S_r^{-1}(u_r)) \\ I_{n3r} &= \widehat{S}_{G_r}(\widehat{S}_r^{-1}(u_r)) - \widehat{S}_{G_r}(S_r^{-1}(u_r)) - \Big[S_{G_r}(\widehat{S}_r^{-1}(u_r)) - S_{G_r}(S_r^{-1}(u_r))\Big]. \end{aligned}$$

Lemma B.2.1 and (C7) imply that $I_{n1r} \to 0$ a.s., while Lemma B.2.2, (C7) and (C4) lead to $I_{n2r} \to 0$ a.s. For I_{n3r} it holds by Lemma B.2.2 and (C7) that

$$|I_{n3r}| \leq \sup_{0 \leq x,y \leq T} \sup_{|x-y| \leq a_n} |[\widehat{S}_{G_r}(x) - \widehat{S}_{G_r}(y)] - [S_{G_r}(x) - S_{G_r}(y)]|$$

for some $T < T_{H_r}$ and $a_n = O(n^{-1/2}(\log n)^{1/2})$. By Lemma B.2.5 the latter is $O(n^{-3/4}(\log n)^{3/4})$ a.s. □

Remark 6.4.1. In practice the copula $\widetilde{C}$ is taken to be of some parametric form, i.e., $\widetilde{C} = \widetilde{C}_\theta$ with known parameter θ. If θ is unknown, then, with $\widehat{\theta}$ a consistent estimator for θ

$$\widetilde{C}_{\widehat{\theta}}(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) \to \widetilde{C}_\theta(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2))) \text{ a.s.}$$

provided the conditions of Theorem 6.4.3 hold and $\widetilde{C}_\theta$ is Lipschitz of order 1 as a function of θ.

The general conclusion of this section is that, under the stated conditions, $\widehat{C}(u_1, u_2) \to C(u_1, u_2)$ a.s. for all (u_1, u_2) in the unit square except (possibly) in strips along the west and the south side. For univariate censoring, the region of convergence is $[S_1(T_{\widetilde{H}}), 1] \times [S_2(T_{\widetilde{H}}), 1]$, while for copula censoring the region is $[S_1(T_{H_1}), 1] \times [S_2(T_{H_2}), 1]$.

6.5 Asymptotic representation

In this section we construct an asymptotic i.i.d. representation for the numerator $\widehat{A}$ and the denominator $\widehat{B}$ of the estimator $\widehat{C}(u_1, u_2)$. Note that

$$\begin{aligned}\widehat{C}(u_1, u_2) - C(u_1, u_2) &= \frac{\widehat{A}}{\widehat{B}} - \frac{A}{B} \\ &= \frac{1}{B}(\widehat{A} - A) - \frac{A}{B^2}(\widehat{B} - B) - \frac{1}{B\widehat{B}}(\widehat{A} - A)(\widehat{B} - B) + \frac{A}{B^2\widehat{B}}(\widehat{B} - B)^2.\end{aligned}$$

The first and second term in the right hand side of the decomposition give an asymptotic i.i.d. representation for the estimator $\widehat{C}(u_1, u_2)$. Indeed, based on the consistency and the theorems in this section, the third and fourth term are $o_P(n^{-1/2})$.

We first give an asymptotic i.i.d. representation for $\widehat{A}$.

Theorem 6.5.1. *Assume (C1), (C2), (C3), (C7) and (C9). Then, as $n \to \infty$,*

$$\begin{aligned}\widehat{A} - A &= \frac{1}{n}\sum_{s=1}^{n}\Big[I(Y_{s1} > S_1^{-1}(u_1), Y_{s2} > S_2^{-1}(u_2)) - S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2)) \\ &\quad + \sum_{r=1}^{2}\frac{1}{S_r'(S_r^{-1}(u_r))}S_Y^{(r)}(S_1^{-1}(u_1), S_2^{-1}(u_2))\ \psi_{sr}(S_r^{-1}(u_r))\Big] + o_P(n^{-1/2})\end{aligned}$$

with $S_Y^{(r)}$, $r = 1, 2$, the partial derivatives of S_Y and

$$\begin{aligned}\psi_{sr}(t) &= S_r(t)\Big[\int_0^t \frac{I(Y_{sr} \le y) - H_r(y)}{(1 - H_r(y))^2}dH_r^u(y) \\ &\quad + \frac{I(Y_{sr} \le t, \delta_{sr} = 1) - H_r^u(t)}{1 - H_r(t)} - \int_0^t \frac{I(Y_{sr} \le y, \delta_{sr} = 1) - H_r^u(y)}{(1 - H_r(y))^2}dH_r(y)\Big]\end{aligned}$$

where $H_r^u(t) = P(Y_r \leq t, \delta_r = 1)$ *(*$r = 1, 2$*).*

Proof. We have

$$\widehat{A} - A = A_{n1} + A_{n2} + A_{n3} + O(n^{-1}) \text{ a.s.}$$

with A_{n1}, A_{n2} and A_{n3} as in the proof of Theorem 6.4.1.

We keep A_{n1} as the first part of the representation.

A Taylor expansion for A_{n3} gives

$$\begin{aligned} A_{n3} &= S_Y^{(1)}(S_1^{-1}(u_1), S_2^{-1}(u_2))(\widehat{S}_1^{-1}(u_1) - S_1^{-1}(u_1)) \\ &+ S_Y^{(2)}(S_1^{-1}(u_1), S_2^{-1}(u_2))(\widehat{S}_2^{-1}(u_2) - S_2^{-1}(u_2)) + o_P(n^{-1/2}). \end{aligned}$$

The order of the remainder term follows from (C3) and Lemma B.2.2. For $\widehat{S}_r^{-1}(u_r) - S_r^{-1}(u_r)$ $(r = 1, 2)$, we use the i.i.d. representation given in Lemma B.2.4:

$$\widehat{S}_r^{-1}(u_r) - S_r^{-1}(u_r) = \frac{1}{nS_r'(S_r^{-1}(u_r))} \sum_{s=1}^{n} \psi_{sr}(S_r^{-1}(u_r)) + R_{nr}$$

where $R_{nr} = O(n^{-1} \log n)$ a.s.

Finally, use Lemma B.2.2 as well as Lemma B.2.6 to obtain that $A_{n2} = O(n^{-3/4}(\log n)^{3/4})$ a.s. □

In Theorem 6.5.2 we obtain, for univariate censoring, an asymptotic i.i.d. representation for $\widehat{B}$. A similar result, for copula censoring, is given in Remark 6.5.1.

Theorem 6.5.2 (univariate censoring)**.** *Assume (C1), (C6), (C8), (C9) and (C11). Then, as* $n \to \infty$*,*

$$\widehat{B} - B =$$

$$\begin{cases} -\frac{1}{n} \sum_{s=1}^{n} \psi_s^G(S_1^{-1}(u_1)) + \dfrac{S_G'(S_1^{-1}(u_1))}{nS_1'(S_1^{-1}(u_1))} \sum_{s=1}^{n} \psi_{s1}(S_1^{-1}(u_1)) + o_P(n^{-1/2}) \\ \qquad\qquad \text{if } S_1^{-1}(u_1) \geq S_2^{-1}(u_2) \\ -\frac{1}{n} \sum_{s=1}^{n} \psi_s^G(S_2^{-1}(u_2)) + \dfrac{S_G'(S_2^{-1}(u_2))}{nS_2'(S_2^{-1}(u_2))} \sum_{s=1}^{n} \psi_{s2}(S_2^{-1}(u_2)) + o_P(n^{-1/2}) \\ \qquad\qquad \text{if } S_2^{-1}(u_2) \geq S_1^{-1}(u_1) \end{cases}$$

with S_G' *the derivative of* S_G *and* $\psi_{sr}(t)$ *(*$r = 1, 2$*) as in Theorem 6.5.1.*

Proof. If $S_1^{-1}(u_1) = S_2^{-1}(u_2)$, then, by consistency of $\widehat{S}_r^{-1}(u_r)$ $(r = 1, 2)$ (Lemma B.2.2) $\widehat{S}_1^{-1}(u_1) = \widehat{S}_2^{-1}(u_2)$ with probability one.

Suppose $S_1^{-1}(u_1) > S_2^{-1}(u_2)$. By (6.6) it holds that $\max(S_1^{-1}(u_1), S_2^{-1}(u_2)) = S_1^{-1}(u_1)$ and $\max(\widehat{S}_1^{-1}(u_1), \widehat{S}_2^{-1}(u_2)) = \widehat{S}_1^{-1}(u_1)$ for n sufficiently large. We then have

$$\widehat{B} - B = B_{n11} + B_{n21} + B_{n31}$$

where

$$\begin{aligned} B_{n11} &= \widehat{S}_G(S_1^{-1}(u_1)) - S_G(S_1^{-1}(u_1)) \\ B_{n21} &= S_G(\widehat{S}_1^{-1}(u_1)) - S_G(S_1^{-1}(u_1)) \\ B_{n31} &= \widehat{S}_G(\widehat{S}_1^{-1}(u_1)) - \widehat{S}_G(S_1^{-1}(u_1)) - S_G(\widehat{S}_1^{-1}(u_1)) + S_G(S_1^{-1}(u_1)). \end{aligned}$$

Under (C1) and (C8) an i.i.d. representation for B_{n11} can be obtained via Lemma B.2.3:

$$B_{n11} = -\frac{1}{n}\sum_{s=1}^{n} \psi_s^G(S_1^{-1}(u_1)) + R_{n11}$$

with $R_{n11} = O(n^{-1}\log n)$ a.s.

A Taylor expansion for B_{n21} gives

$$S_G(\widehat{S}_1^{-1}(u_1)) - S_G(S_1^{-1}(u_1)) = S_G'(S_1^{-1}(u_1))(\widehat{S}_1^{-1}(u_1) - S_1^{-1}(u_1)) + o_P(n^{-1/2}).$$

The order of the remainder term follows from (C6) and Lemma B.2.2. For $\widehat{S}_1^{-1}(u_1) - S_1^{-1}(u_1)$ we use the i.i.d. representation given in Lemma B.2.4:

$$\widehat{S}_1^{-1}(u_1) - S_1^{-1}(u_1) = \frac{1}{n}\frac{1}{S_1'(S_1^{-1}(u_1))}\sum_{s=1}^{n} \psi_{s1}(S_1^{-1}(u_1)) + R_{n1}$$

where $R_{n1} = O(n^{-1}\log n)$ a.s.

Further, Lemma B.2.2 and Lemma B.2.5 imply that $B_{n31} = O(n^{-3/4}(\log n)^{3/4})$ a.s. The proof is analogous for $S_2^{-1}(u_2) > S_1^{-1}(u_1)$. □

Remark 6.5.1. Since

$$\begin{aligned} & \widetilde{C}(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) = \widetilde{C}(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2))) \\ + \;& \widetilde{C}^{(1)}(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2)))[\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)) - S_{G_1}(S_1^{-1}(u_1))] \\ + \;& \widetilde{C}^{(2)}(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2)))[\widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2)) - S_{G_2}(S_2^{-1}(u_2))] \\ + \;& o_P(n^{-1/2}) \end{aligned}$$

the i.i.d. representation for $\widehat{B} - B$ in case of copula censoring can be obtained along the same lines if we assume that $\widetilde{C}$ has Lipschitz continuous partial derivatives $\widetilde{C}^{(1)}$ and $\widetilde{C}^{(2)}$. In practice the copula $\widetilde{C}$ is taken to be of some parametric form, i.e., $\widetilde{C} = \widetilde{C}_\theta$ with known parameter θ. If θ is unknown and estimated by $\widehat{\theta}$, we can use the above together with

$$\begin{aligned} & \widetilde{C}_{\widehat{\theta}}(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) = \widetilde{C}_\theta(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) \\ + & \left[\widetilde{C}_{\widehat{\theta}}(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) - \widetilde{C}_\theta(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2)))\right] \end{aligned}$$

as well as

$$\begin{aligned} & \widetilde{C}_{\widehat{\theta}}(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) - \widetilde{C}_\theta(\widehat{S}_{G_1}(\widehat{S}_1^{-1}(u_1)), \widehat{S}_{G_2}(\widehat{S}_2^{-1}(u_2))) \\ = & \widetilde{C}'_\theta(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2)))(\widehat{\theta} - \theta) + o_P(n^{-1/2}) \end{aligned}$$

if we further require that $\widetilde{C}'_\theta$, the partial derivative of $\widetilde{C}_\theta$ with respect to θ, is Lipschitz in θ and in $(S_{G_1}(S_1^{-1}(u_1)), S_{G_2}(S_2^{-1}(u_2)))$ and that $\widehat{\theta} - \theta$ has some i.i.d. representation.

The general conclusion of this section is that, under the stated conditions, $\widehat{C}(u_1, u_2) - C(u_1, u_2)$ can be represented as an average of i.i.d. random variables with zero mean and a term of lower order. The representation is valid in a subset of the unit square, $[S_1(T_{\widetilde{H}}), 1] \times [S_2(T_{\widetilde{H}}), 1]$, resp. $[S_1(T_{H_1}), 1] \times [S_2(T_{H_2}), 1]$ for univariate, resp. copula censoring. Normality can be obtained, but given the complexity of the representations, the derivation of a closed formula for the asymptotic variances is not tractable. Moreover, the latter will contain several expressions that need to be estimated. Although we do not investigate this here, bootstrap alternatives can be considered.

6.6 Simulation study

To evaluate the finite sample behavior of the proposed copula estimator (6.5), we set up a simulation study. A comparison with the estimator of Gribkova and Lopez (2014) in (6.3) is made.

6.6.1 Simulation setting

We generate 500 datasets, each containing 250 or 500 clusters of size 2 from a Clayton copula, resp. a Gumbel copula, with a dependence parameter such that Kendall's tau (τ_T) equals 0.25 or 0.75 (Table 6.1). For the event times T_{sr} ($s = 1, \ldots, 250$ or 500, $r = 1, 2$), Weibull margins with scale $\lambda = 0.5$ and shape $\rho = 1.5$ are used.

Censoring is assumed to be univariate or to be directed by a Clayton copula, resp.

a Gumbel copula, with a dependence parameter such that Kendall's tau ($\tau_{\widetilde{C}}$) equals 0.1 or 0.5 (Table 6.1). For the censoring times C_{sr} ($s = 1, \ldots, 250$ or 500, $r = 1, 2$), we use Weibull margins with scale and shape given by $\lambda = 0.15, \rho = 1.5$, resp. $\lambda = 0.85, \rho = 1.5$, leading to approximately 23%, resp. 63% censoring.

The observed data are $Y_{sr} = \min(T_{sr}, C_{sr})$ and $\delta_{sr} = I(T_{sr} \leq C_{sr})$ ($s = 1, \ldots, 250$ or 500, $r = 1, 2$).

Here, the focus is on a Clayton copula and a Gumbel copula. Even though both copulas can exhibit the same strength of association as expressed by Kendall's tau (τ), they model a different correlation structure, i.e., a Clayton copula is lower tail dependent, while a Gumbel copula is upper tail dependent (Section 5.7 - Table 6.1).

6.6.2 Simulation results

To investigate the performance of the proposed copula estimators as compared to their Gribkova-Lopez equivalent, we compute for each estimator the distance between the estimated copula $\widehat{C}$ and the true underlying copula C, i.e., we look at

$$\text{RMSE} = \sqrt{\sum_{(u_l, u_k) \in \text{Grid}} \left\{ \widehat{C}(u_l, u_k) - C(u_l, u_k) \right\}^2}$$

Table 6.1: Simulation setting. Copula parameter θ with corresponding value of Kendall's tau (τ) and lower (δ_L), resp. upper (δ_U) tail dependence parameter.

copula	θ	τ	δ_L	δ_U
Clayton	0.22	0.10	0.04	0
	0.66	0.25	0.35	0
	2	0.50	0.71	0
	6	0.75	0.89	0
Gumbel	1.11	0.10	0	0.13
	1.33	0.25	0	0.32
	2	0.50	0	0.59
	4	0.75	0	0.81

where Grid $= R \cap \{(u,v)|u,v \in \{0.05, 0.10, \ldots, 0.95\}\}$ with $R = [S_1(T_{\widetilde{H}}), 1] \times [S_2(T_{\widetilde{H}}), 1]$ for univariate censoring and $R = [S_1(T_{H_1}), 1] \times [S_2(T_{H_2}), 1]$ for copula censoring.

Figures 6.1 - 6.5 display boxplots of the difference in RMSE between the proposed and the Gribkova-Lopez estimator (RMSE_P-RMSE_{GL}) as obtained in the diverse settings described in Section 6.6.1. Herein, the wiskers of each boxplot extend to the 5 and 95 percentiles of the obtained RMSE difference. If the true copula is Clayton, resp. Gumbel, the boxplot is labeled by Cijk, resp. Gijk. The indexing is as follows: $i = 1$ for $\tau_T = 0.25$ and $i = 2$ for $\tau_T = 0.75$; $j = 1$ for 23% censoring and $j = 2$ for 63% censoring; $k = 1$ for 250 clusters and $k = 2$ for 500 clusters, e.g., C121 refers to a Clayton copula with $\tau_T = 0.25$, 63% of censoring and 250 clusters. Note that a value of RMSE_P-$\text{RMSE}_{GL} < 0$ favors our estimator over the one of Gribkova and Lopez (2014).

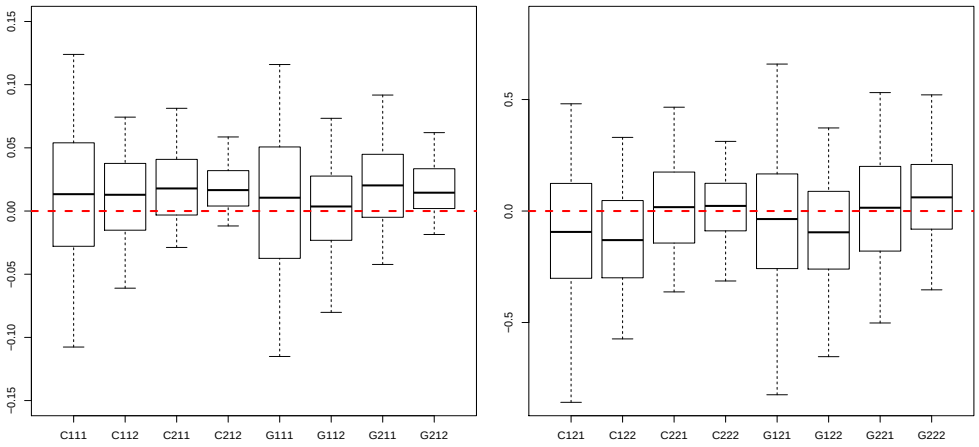
Figures 6.1 (a) and (b) summarize the results in case of univariate censoring. The boxplots indicate that for light censored data, the Gribkova-Lopez estimator outperforms the proposed one, whereas for heavily censored data the proposed estimator performs at least as well and often better than the Gribkova-Lopez estimator, especially when the correlation between event times is low ($\tau_T = 0.25$). In Figures 6.1 (c) and (d) a four-dimensional version of the proposed estimator is compared with the four-dimensional Gribkova-Lopez estimator. Clearly our estimator gains a lot in performance as compared to the Gribkova-Lopez estimator and is the better one.

The results of copula censoring with a known parameter (Figure 6.2 and Figure 6.3) are similar to those where the copula parameter is estimated (Figure 6.4 and Figure 6.5). The boxplots indicate that for light censored data and event times with an association of $\tau_T = 0.25$ both nonparametric estimators have similar performance, while for highly related event times ($\tau_T = 0.75$) the estimator of Gribkova and Lopez is the better one. For heavily censored data the boxplots reveal that for $\tau_T = 0.75$ and $\tau_{\widetilde{C}} = 0.10$ the Gribkova-Lopez estimator is to be preferred, while for $\tau_T = 0.25$ and $\tau_{\widetilde{C}} = 0.50$ the proposed estimator is superior. For the remaining combinations of τ_T and $\tau_{\widetilde{C}}$ no real preference is shown.

These results are in line with what should be expected. The Gribkova-Lopez estimator takes only those clusters into account with all observations events and therefore heavy censoring and/or a high cluster dimension eliminates a considerable amount of data, whereas the proposed copula estimator does not discard this information. We also see that for strongly related event times this elimination is less severe.

(a) univariate censoring, two dimensions

(b) univariate censoring, two dimensions



(c) univariate censoring, four dimensions

(d) univariate censoring, four dimensions

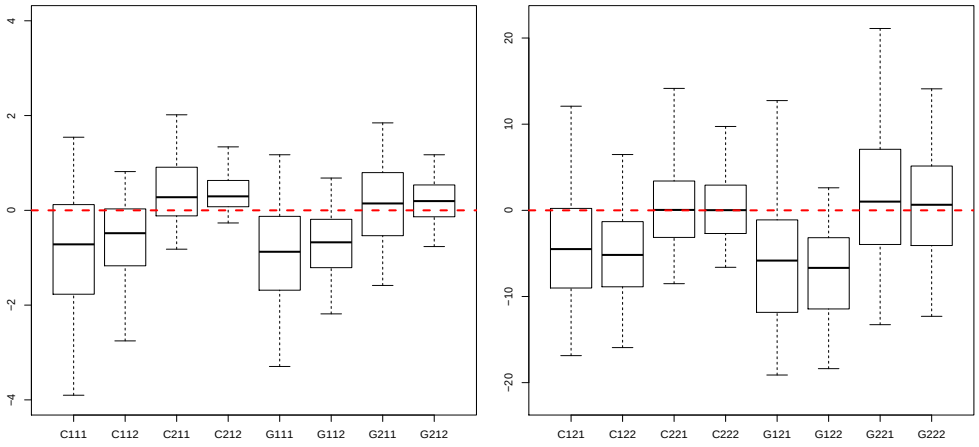
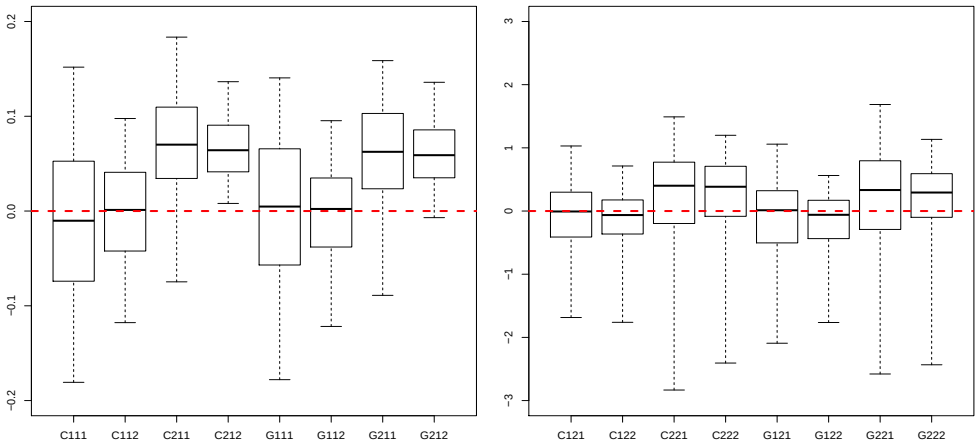


Figure 6.1: Simulation results. Difference in RMSE (RMSE_P-RMSE_{GL}) for univariate censoring. Left panel: 23% censoring, right panel: 63% censoring - upper panel: two-dimensional data, lower panel: four-dimensional data.

(a) censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.1$ **(b)** censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.1$



(c) censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.5$ **(d)** censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.5$

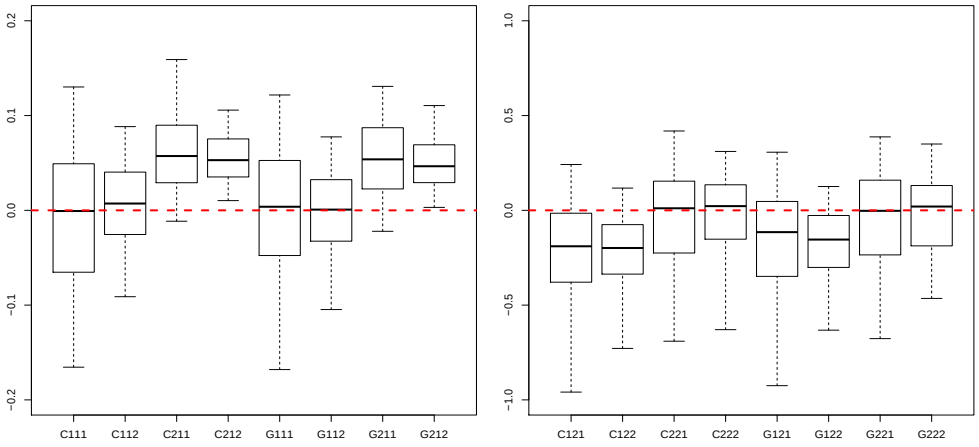
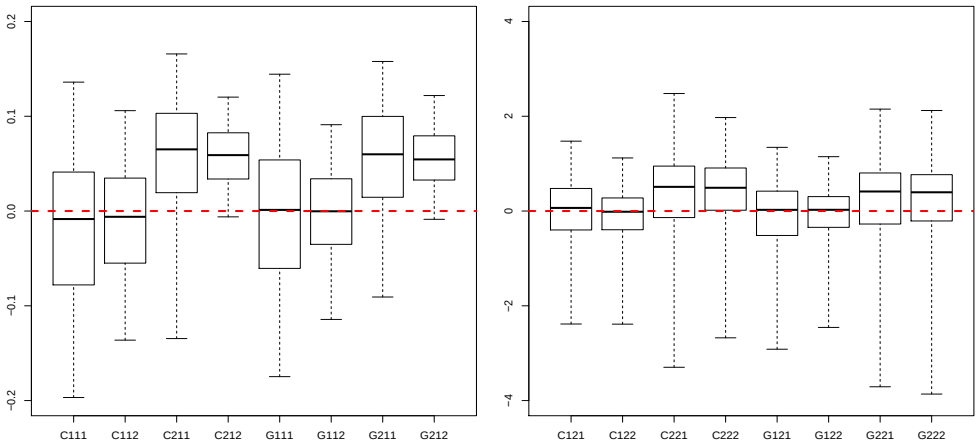


Figure 6.2: Simulation results. Difference in RMSE (RMSE$_P$-RMSE$_{GL}$) for censoring via Clayton copula. Left panel: 23% censoring, right panel: 63% censoring - upper panel: censoring copula has $\tau_{\widetilde{C}} = 0.10$, lower panel: censoring copula has $\tau_{\widetilde{C}} = 0.50$.

(a) censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.1$ **(b)** censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.1$



(c) censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.5$ **(d)** censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.5$

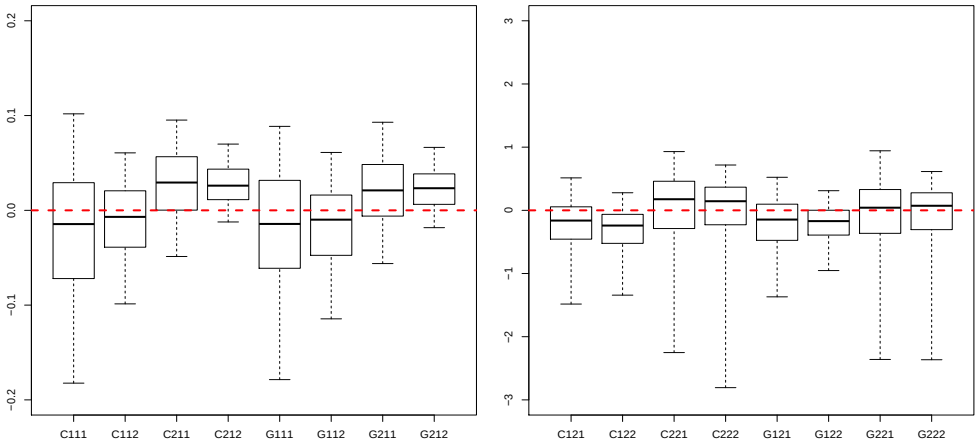
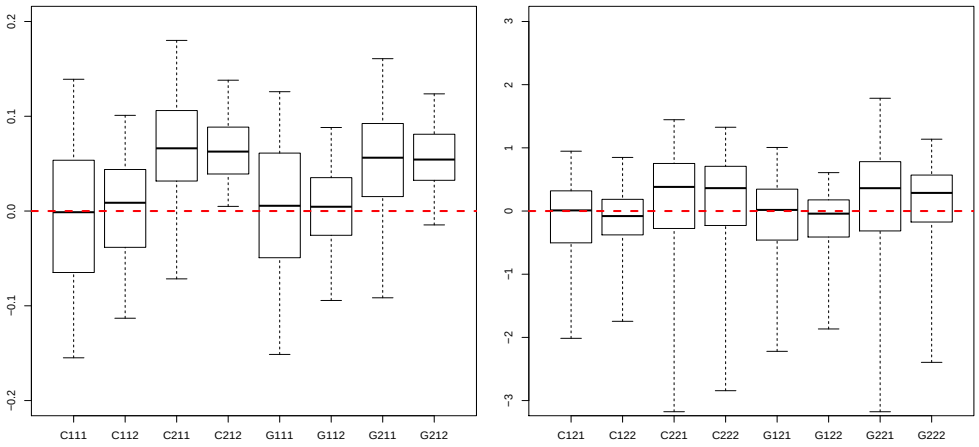


Figure 6.3: Simulation results. Difference in RMSE (RMSE_P-RMSE_{GL}) for censoring via Gumbel copula. Left panel: 23% censoring, right panel: 63% censoring - upper panel: censoring copula has $\tau_{\widetilde{C}} = 0.10$, lower panel: censoring copula has $\tau_{\widetilde{C}} = 0.50$.

(a) censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.1$ **(b)** censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.1$



(c) censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.5$ **(d)** censoring via Clayton copula with $\tau_{\widetilde{C}} = 0.5$

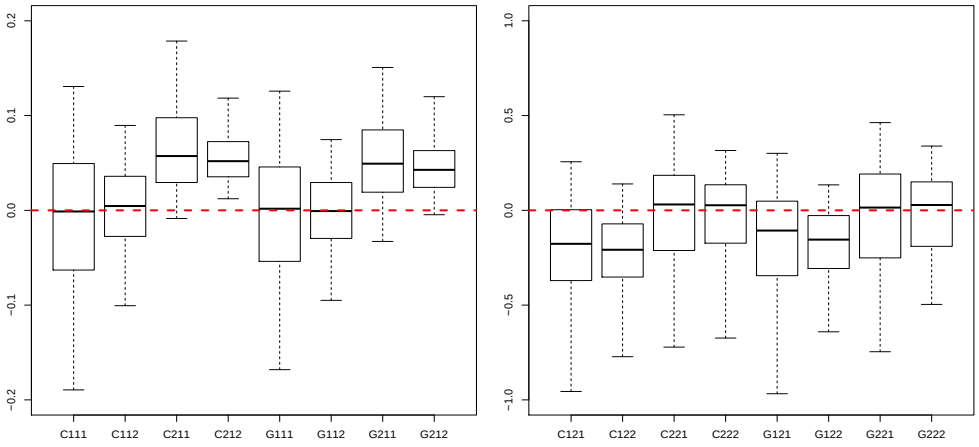
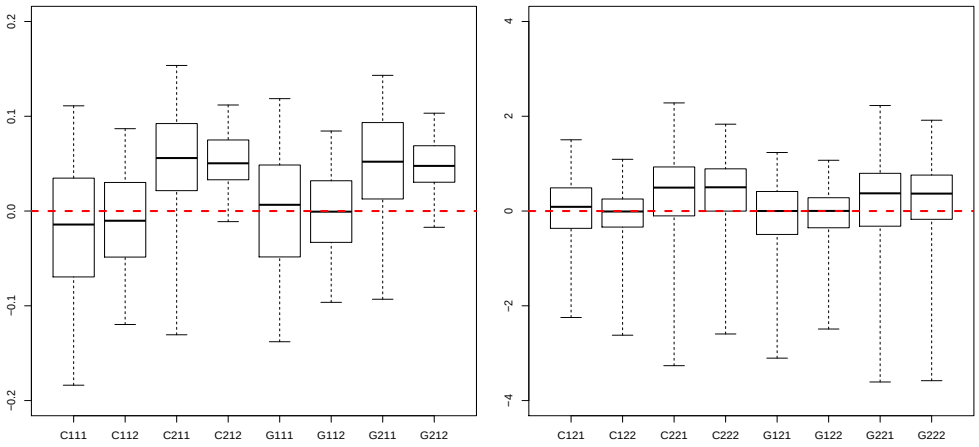


Figure 6.4: Simulation results. Difference in RMSE (RMSE$_P$-RMSE$_{GL}$) for censoring via estimated Clayton copula. Left panel: 23% censoring, right panel: 63% censoring - upper panel: censoring copula has $\tau_{\widetilde{C}} = 0.10$, lower panel: censoring copula has $\tau_{\widetilde{C}} = 0.50$.

(a) censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.1$ **(b)** censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.1$



(c) censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.5$ **(d)** censoring via Gumbel copula with $\tau_{\widetilde{C}} = 0.5$

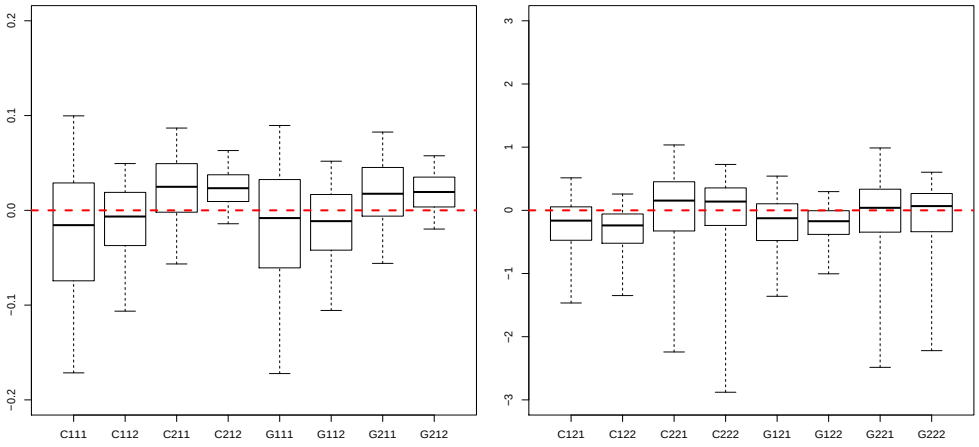


Figure 6.5: Simulation results. Difference in RMSE (RMSE_P-RMSE_{GL}) for censoring via estimated Gumbel copula. Left panel: 23% censoring, right panel: 63% censoring - upper panel: censoring copula has $\tau_{\widetilde{C}} = 0.10$, lower panel: censoring copula has $\tau_{\widetilde{C}} = 0.50$.

6.7 Discussion

In this chapter, we define a new nonparametric copula estimator for the joint survival function of right-censored event time data. Censoring is assumed to be univariate or to be dictated by a copula. We establish consistency and obtain an asymptotic i.i.d. representation for the proposed estimator. A simulation study reveals that, depending on the amount of censoring, the strength of data association and the data dimension either our estimator or its Gribkova-Lopez equivalent is preferred.

Even though the new nonparametric copula estimator covers a wide range of right-censored data settings, one may want to relax the assumption of univariate or copula censoring. To this end, the nonparametric survival function estimator by Akritas and Van Keilegom (2003) might be worthwhile to investigate. The latter is a linear combination of averaged kernel based conditional survival functions, which is shown to be consistent and to exhibit asymptotic normality. Via Sklar's theorem and an appropriate choice of marginal quantile functions, a nonparametric copula estimator can be constructed. However, compared to the proposed nonparametric copula estimator, bandwidth selection needs to be handled.

Often a semiparametric copula is used to analyze right-censored grouped event time data (Chapter 5). Given that the copula determines the type of association between the cluster components, it is essential to verify the aptness of the chosen copula. The newly derived nonparametric copula estimator can be used to do so, e.g., via the L_2-distance with the maximum quasi-loglikelihood estimator of the presumed copula (Section 5.2). A comparison with copula selection via AIC and/or BIC (Section 5.3) can be made. This topic is the subject of future research.

6.8 Addendum

Nonparametric survival function estimation has also been considered by van der Laan *et al.* (2002). They focus on univariate right-censored event time data. In this section we extend the van der Laan estimator to general right-censored event time data and we show that, for event times and censoring times that are independent, the estimator in van der Laan *et al.* (2002) reduces to estimator (6.2).

For cluster s, define $B_s = I(T_{s1} > t_1, T_{s2} > t_2)$ and $\Delta_s = \prod_{r=1}^{2}(1 - \Delta_{sr})$ with $\Delta_{sr} = I(\min(T_{sr}, t_r) > C_{sr})$ ($s = 1, \ldots, n$ and $r = 1, 2$).

Following the idea in van der Laan *et al.* (2002), we have: B_s can not be calculated from the observed data (Y_{sr}, δ_{sr}) ($s = 1, \ldots, n$ and $r = 1, 2$)

$$\begin{aligned} &\Leftrightarrow \quad \exists r \in \{1,2\} : \delta_{sr} = 0 \text{ and } t_r > C_{sr} \\ &\Leftrightarrow \quad \exists r \in \{1,2\} : T_{sr} > C_{sr} \text{ and } t_r > C_{sr} \\ &\Leftrightarrow \quad \exists r \in \{1,2\} : \min(T_{sr}, t_r) > C_{sr} \\ &\Leftrightarrow \quad \exists r \in \{1,2\} : \Delta_{sr} = 1 \end{aligned}$$

or equivalently, B_s can be calculated from the observed data $(Y_{sr}, \delta_{sr}) \Leftrightarrow \Delta_s = 1$ ($s = 1, \ldots, n$ and $r = 1, 2$).

Further,

$$\begin{aligned} \mathsf{E}[\Delta | T_1, T_2] &= \mathsf{E}\big[\textstyle\prod_{r=1}^{2}(1 - \Delta_r) | T_1, T_2\big] \\ &= \mathsf{E}[I(\min(T_1, t_1) \le C_1, \min(T_2, t_2) \le C_2) | T_1, T_2] \\ &= P(\min(T_1, t_1) \le C_1, \min(T_2, t_2) \le C_2 \ | T_1, T_2) \\ &= S_C^{lc}(\min(T_1, t_1), \min(T_2, t_2) | T_1, T_2) \end{aligned}$$

and thus

$$\begin{aligned} & \mathsf{E}\left[\frac{\Delta\, B}{S_C^{lc}(\min(T_1, t_1), \min(T_2, t_2) | T_1, T_2)}\right] \\ = \ & \mathsf{E}\left[\mathsf{E}\left[\frac{\Delta\, B}{S_C^{lc}(\min(T_1, t_1), \min(T_2, t_2) | T_1, T_2)}\right] | T_1, T_2\right] \\ = \ & \mathsf{E}\left[\frac{B\, \mathsf{E}\Big[\Delta | T_1, T_2\Big]}{S_C^{lc}(\min(T_1, t_1), \min(T_2, t_2) | T_1, T_2)}\right] = \mathsf{E}[B]. \end{aligned}$$

Based on the above, an estimator is given by

$$\widetilde{S}(t_1, t_2) = \frac{1}{n} \sum_{s=1}^{n} \frac{\Delta_s\, B_s}{\widehat{S}_C^{lc}(\min(T_{s1}, t_1), \min(T_{s2}, t_2) | T_{s1}, T_{s2})} \tag{6.7}$$

with $\widehat{S}_C^{lc}$ an appropriate estimator for a left-continuous version of S_C.

To see the equivalence with estimator (6.2), note that $I(Y_{s1} > t_1, Y_{s2} > t_2) = \Delta_s\, B_s$. Further, the denominator of (6.7) needs to be calculated only if $\Delta_s\, B_s = 1$. The latter implies $T_{sr} > t_r$ ($r = 1, 2$) and we obtain $\widehat{S}_C^{lc}(t_1, t_2 | T_{s1}, T_{s2})$. For event times and censoring times that are independent, we have $\widehat{S}_C^{lc}(t_1, t_2)$. Hence, (6.2) and (6.7) are equivalent, except for the use of a right-continuous, resp. a left-continuous estimator for S_C.

Bibliography

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009) Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, **44**, 182–198.

Aerts, M., Claeskens, G. and Hart, J. D. (1999) Testing the fit of a parametric function. *Journal of the American Statistical Association*, **94**, 869–879.

Aerts, M., Claeskens, G. and Hart, J. D. (2004) Bayesian-motivated tests of function fit and their asymptotic frequentist properties. *The Annals of Statistics*, **32**, 2580–2615.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory* (Eds. B. Petrov and F. Csáki), 267–281. Akadémiai Kiadó, Budapest.

Akritas, M. G. and Van Keilegom, I. (2003) Estimation of bivariate and marginal distributions with censored data. *Journal of the Royal Statistical Society: Series B*, **65**, 457–471.

Aly, E., Csörgő, M. and Horváth, L. (1985) Strong approximations of the quantile process of the product-limit estimator. *Journal of Multivariate Analysis*, **16**, 185–210.

Bahadur, R. (1966) A note on quantiles in large samples. *The Annals of Mathematical Statistics*, **37**, 577–580.

Berg, D. and Aas, K. (2009) Models for construction of multivariate dependence: A comparison study. *The European Journal of Finance*, **15**, 639–659.

Chen, X., Fan, Y., Pouzo, D. and Ying, Z. (2010) Estimation and model selection of semiparametric multivariate survival functions under general censorship. *Journal of Econometrics*, **157**, 129–142.

Claeskens, G. and Hart, J. D. (2009) Goodness-of-fit tests in mixed models. *Test*, **18**, 213–239. With discussion.

Claeskens, G. and Hjort, N. L. (2008) *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

Cui, S. and Sun, Y. (2004) Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. *Statistica Sinica*, **14**, 249–267.

Davison, A. and Hinkley, D. (1997) *Bootstrap methods and their applications.* Cambridge University Press.

Deheuvels, P. (1979) La fonction de dépendance empirique et ses propriétés - un test non paramétrique d'indépendance. *Académie Royale de Belgique - Bulletin de la Classe des Sciences - 5e Série*, **65**, 274–292.

Duchateau, L. and Janssen, P. (2004) Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. *Biometrics*, **60**, 608–614.

Duchateau, L. and Janssen, P. (2008) *The Frailty Model*. Springer, New York.

Duchateau, L., Opsomer, G., Dewulf, J. and Janssen, P. (2005) The nonlinear effect (determined by the penalised partial likelihood approach) of milk-protein concentration on time to first insemination in belgian dairy cows. *Preventive Veterinary Medicine*, **68**, 81–90.

Fermanian, J., Radulovic, D. and Wegkamp, M. (2004) Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847–860.

Földes, A. and Rejtő, L. (1981) A lil type result for the product-limit estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **56**, 75–86.

Gaenssler, P. and Stute, W. (1987) Seminar on empirical processes. *DMV Sem. 9. Basel: Birkhaüser*.

Geerdens, C., Claeskens, G. and Janssen, P. (2013) Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics*, **14**, 433–446.

Geerdens, C., Claeskens, G. and Janssen, P. (2014) Copula based flexible modeling of associations between clustered event times. *Technical report, submitted for publication*.

Geerdens, C., Janssen, P. and Veraverbeke, N. (2015) Large sample properties of non-parametric copula estimators under bivariate censoring. *Technical report, submitted for publication*.

Gijbels, I. (1990) Asymptotic representations under random censoring. *Unpublished PhD-thesis, Limburgs Universitair Centrum, Belgium*.

Gijbels, I. and Veraverbeke, N. (1998) Weak asymptotic representations for quantiles of the product-limit estimator. *Journal of Statistical Planning and Inference*, **18**, 151–160.

Glidden, D. V. (1999) Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*, **86**, 381–393.

Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and Duchateau, L. (2009) Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 1–14.

Goethals, K., Janssen, P. and Duchateau, L. (2008) Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, **35**, 1071–1079.

Gribkova, S. and Lopez, O. (2014) Nonparametric copula estimation under bivariate censoring. *Technical report*.

Gribkova, S., Lopez, O. and Saint-Pierre, P. (2013) A simplified model for studying bivariate mortality under right-censoring. *Journal of Multivariate Analysis*, **115**, 181–192.

Grønneberg, S. and Hjort, N. L. (2014) The copula information criteria. *The Scandinavian Journal of Statistics*, **41**, 436–459.

Hofert, M. (2008) Sampling Archimedean copulas. *Computational Statistics and Data Analysis*, **52**, 5163–5174.

Joe, H. (1993) Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, **46**, 262–282.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall.

Joe, H. and Hu, T. (1996) Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, **57**, 240–265.

Kaplan, E. and Meier, P. (1958) Nonparametric estimation of incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Klein, J. and Moeschberger, M. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. Springer.

Kurowicka, D. and Joe, H. e. (2011) *Dependence modeling: vine copula handbook*. World Scientific.

van der Laan, M., Hubbard, A. and Robins, J. (2002) Locally efficient estimation of a multivariate survival function in longitudinal studies. *Journal of the American Statistical Association*, **97**, 494–507.

Lin, D. and Ying, Z. (1993) A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, **80**, 573–581.

Lo, S. and Singh, K. (1986) The product-limit estimator and the bootstrap: some asymptotic representations. *Probability Theory and Related Fields*, **71**, 455–465.

Lopez, O. and Saint-Pierre, P. (2012) Bivariate censored regression relying on a new estimator of the joint distribution function. *Journal of Statistical Planning and Inference*, **142**, 2440–2453.

Maechler, M. (2014) *Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable*. URL `http://CRAN.R-project.org/package=Rmpfr`. R package version 0.5-6.

Major, P. and Retjő, L. (1988) Strong embedding of the estimator of the distribution function under random censorship. *The Annals of Statistics*, **16**, 1113–1132.

Massonnet, G., Burzykowski, T. and Janssen, P. (2006) Resampling plans for frailty models. *Communications in Statistics-Simulation and Computation*, **35**, 497–514.

Massonnet, G., Janssen, P. and Duchateau, L. (2009) Modelling udder infection data using copula models for quadruples. *Journal of Statistical Planning and Inference*, **139**, 3865–3877.

Nelsen, R. B. (2006) *An introduction to copulas*. Springer.

Nikiforov, A. F. and Uvarov, V. B. (1988) *Special functions of mathematical physics*. Basel: Birkhäuser Verlag. A unified introduction with applications, Translated from the Russian and with a preface by Ralph P. Boas, With a foreword by A. A. Samarskiĭ.

Savu, C. and Trede, M. (2006) Hierarchical Archimedean copulas. *International Conference on High Frequency Finance, Konstanz, Germany*.

Schäfer, H. (1986) Local convergence of empirical measures in the random censorship situation with application to density and rate estimators. *The Annals of Statistics*, **14**, 1240–1245.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Segers, J. (2012) Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, **18**, 764–782.

Serfling, R. (1980) *Approximation theorems of mathematical statistics*. Wiley.

Shih, J. H. and Louis, T. A. (1995) Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis.*, **1**, 205–220.

Sin, C. and White, H. (1996) Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, **71**, 207–225.

Sklar, M. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l' Intitut de Statistique de l' Université de Paris*, **8**, 229–231.

Thas, O. (2009) Comments on: Goodness-of-fit tests in mixed models. *Test*, **18**, 260–264.

Therneau, T. and Grambsch, P. (2000) *Modeling Survival Data. Extending the Cox Model*. Springer, New York.

Wang, W. and Wells, M. (1997) Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, **84**, 863–880.

Wienke, A. (2011) *Frailty Models in Survival Analysis*. Chapman-Hall.

Zhang, D. and Davidian, M. (2001) Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**, 795–802.

Appendix A

Bootstrap algorithms

A.1 Frailty bootstrap algorithm

For the order selection test in Section 3.3, an approximate p-value can be obtained via a parametric bootstrap as described in Algorithms 1 and 2 (Davison and Hinkley, 1997; Massonnet *et al.*, 2006).

Algorithm 1

Step 1: Fit the loglikelihood in Theorem 3.2.1 with $m = 0, \ldots, M_n$ and obtain the actual value of the order selection test statistic: T_{act}. Denote the parameter estimates under the null hypothesis ($m = 0$) by $\widehat{\xi}_0$, $\widehat{\beta}_0$ and $\widehat{\theta}_0$.

Step 2: Generate B resamples in the following way:

Step 2.1: Sample $u_1^*, \ldots, u_S^*$ from a $\Gamma(1/\widehat{\theta}_0, \widehat{\theta}_0)$ distribution.

Step 2.2: For $r = 1, \ldots, n_s$ and $s = 1, \ldots, n$, generate event times T_{sr}^* from the estimated survival function $\widehat{S}_{sr}(t) = \{\exp(-\widehat{H}_0(t))\}^{u_s^* \exp(\widehat{\beta}_0 x_{sr})}$ with $\widehat{H}_0(t)$ the estimated cumulative baseline hazard.

Step 2.3: If $\delta_{sr} = 0$ set $C_{sr}^* = Y_{sr}$; if $\delta_{sr} = 1$ generate C_{sr}^* from the same uniform distribution as was used for the original data.

Step 2.4: Set $Y_{sr}^* = \min(T_{sr}^*, C_{sr}^*)$ with $\delta_{sr}^* = 1$ if $Y_{sr}^* = T_{sr}^*$; and $\delta_{sr}^* = 0$ otherwise.

Step 2.5: Obtain the bootstrap value of the order selection test statistic: T_b^*.

Step 3: Obtain the bootstrap version of $P_{\mathcal{H}_0}(T > T_{act})$, i.e., $p^* = \#\{b : T_b^* > T_{act}\}/B$.

Algorithm 2

This algorithm is analogous to algorithm 1, with Step 2.3 replaced by:

Step 2.3 bis: If $\delta_{sr} = 0$ set $C_{sr}^* = Y_{sr}$; if $\delta_{sr} = 1$ generate C_{sr}^* from the conditional censoring distribution given that $C_{sr} > Y_{sr}$, i.e., generate C_{rs}^* from

$$\frac{\widehat{G}(t) - \widehat{G}(Y_{sr})}{1 - \widehat{G}(Y_{sr})}$$

with $\widehat{G}$ the Kaplan-Meier estimator of the censoring distribution.

A.2 Copula bootstrap algorithm

In the presence of univariate censoring, standard errors for the estimated parameters of a copula can be obtained via a parametric bootstrap as described in Algorithms 1 and 2 (Davison and Hinkley, 1997; Massonnet *et al.*, 2009).

Algorithm 1

Step 1: Fit the copula of interest to the data $(\widehat{V}_{sr}, \delta_{sr})$, where $Y_{sr} = \min(T_{sr}, C_s)$, $\delta_{sr} = I(T_{sr} \leq C_s)$ and $\widehat{V}_{sr} = \widehat{S}_r(Y_{sr})$ with $\widehat{S}_r$ the Kaplan-Meier estimator for the true survival function of the observations T_{sr} ($s = 1, \ldots, n$ and $r = 1, \ldots, 4$). Obtain the maximum quasi-likelihood estimator $\widehat{\zeta}_{S,d}$ (Section 5.2).

Step 2: Generate B bootstrap resamples in the following way:

Step 2.1: Generate $(U_{s1}^*, U_{s2}^*, U_{s3}^*, U_{s4}^*)$ from the copula of interest with $\widehat{\zeta}_{S,d}$ as parameter value.

Step 2.2: Create $(T_{s1}^*, T_{s2}^*, T_{s3}^*, T_{s4}^*)$ via $T_{sr}^* = \widehat{S}_r^{-1}(U_{sr}^*)$.

Step 2.3: Estimate the censoring distribution G via a Kaplan-Meier estimator based on the observations $(\max(Y_{s1}, Y_{s2}, Y_{s3}, Y_{s4}), 1 - \delta_{s1}\delta_{s2}\delta_{s3}\delta_{s4})$. Generate C_s^* from $\widehat{G}$.

Step 2.4: Set $Y_{sr}^* = \min(T_{sr}^*, C_s^*)$ and $\delta_{sr}^* = I(T_{sr}^* \leq C_s^*)$.

Step 2.5: Set $\widehat{V}_{sr}^* = \widehat{S}_r^*(Y_{sr}^*)$ with $\widehat{S}_r^*$ the Kaplan-Meier estimator for the true survival function of the observations T_{sr}^*.

Step 2.6: Fit the copula of interest to the bootstrap data $(\widehat{V}^*_{sr}, \delta^*_{sr})$. Obtain the maximum quasi-likelihood estimator $\widehat{\zeta}^*_{S,d}$.

Step 3: The B bootstrap resamples give $\widehat{\zeta}^{*(1)}_{S,d}, \ldots, \widehat{\zeta}^{*(B)}_{S,d}$. Calculate the standard error of the $\widehat{\zeta}^{*(b)}_{S,d}$ $(b = 1, \ldots, B)$.

Algorithm 2

This algorithm is analogous to algorithm 1, with Step 2.3 replaced by:

Step 2.3 bis: If $\delta_{sr} = 0$ for at least one $r \in \{1, 2, 3, 4\}$, set $C^*_s = \max(Y_{s1}, Y_{s2}, Y_{s3}, Y_{s4})$; if $\delta_{sr} = 1$ for all $r \in \{1, 2, 3, 4\}$ generate C^*_s from the conditional censoring distribution given that $C_s > Y_{sr}$, i.e., generate C^*_s from

$$\frac{\widehat{G}(t) - \widehat{G}(Y_{sr})}{1 - \widehat{G}(Y_{sr})}$$

with $\widehat{G}$ the Kaplan-Meier estimator of the censoring distribution based on the observations $(\max(Y_{s1}, Y_{s2}, Y_{s3}, Y_{s4}), 1 - \delta_{s1}\delta_{s2}\delta_{s3}\delta_{s4})$.

Appendix B

Useful propositions, theorems and lemmas

B.1 Propositions and theorems used in Chapter 5

B.1.1 Propositions used in Section 5.3

The proofs of the propositions in Section 5.3 rely on results in Chen *et al.* (2010). In this section we list the required propositions and their conditions.

Consider four-variate right-censored event time data. Denote the observed time of item r $(r = 1, \ldots, 4)$ in cluster s $(s = 1, \ldots, n)$ by $Y_{sr} = \min(T_{sr}, C_{sr})$ with T_{sr} the true event time and C_{sr} the censoring time. The indicator $\delta_{sr} = I(T_{sr} \leq C_{sr})$ equals one if $Y_{sr} = T_{sr}$ and zero otherwise. With S_r the survival function of the rth component and $\widehat{S}_r$ the corresponding Kaplan-Meier estimator, define $V_{sr} = S_r(Y_{sr})$ and $\widehat{V}_{sr} = \widehat{S}_r(Y_{sr})$ $(s = 1, \ldots, n$ and $r = 1, \ldots, 4)$.

The following conditions are sufficient to ensure the convergence of the maximum quasi-loglikelihood estimator $\widehat{\zeta}_{n,d}$ to the pseudo-true parameter value $\zeta^*_{n,d}$ (Chen *et al.*, 2010).

(C1) (i) The sequence of event times $(T_{s1}, \ldots, T_{s4})$ $(s = 1, \ldots, n)$ is an i.i.d. sample from an unknown joint survival function S with continuous marginal survival functions S_r $(r = 1, \ldots, 4)$.

(ii) The sequence of censoring times $(C_{s1}, \ldots, C_{s4})$ $(s = 1, \ldots, n)$ is an inde-

pendent sample with joint survival functions $S_{C_s}(c_1, \ldots, c_4) = P(C_{s1} > c_1, \ldots, C_{s4} > c_4)$ $(s = 1, \ldots, n)$ and marginal survival functions $S_{G_{sr}}$ $(s = 1, \ldots, n$ and $r = 1, \ldots, 4)$.

(iii) The censoring times $(C_{s1}, \ldots, C_{s4})$ are independent of the event times $(T_{s1}, \ldots, T_{s4})$ $(s = 1, \ldots, n)$ and there is no mass concentration at 0 in the sense that $\limsup_{n\to\infty} n^{-1} \sum_{s=1}^{n} (1 - S_{G_{sr}}(\eta)) \to 0$ as $\eta \to 0$ $(r = 1, \ldots, 4)$.

(C2) Let A_d, the parameter space of ζ_d, be a compact subset of $\mathbb{R}^{q_d}$. For every $\epsilon > 0$,

$$\liminf_{\zeta_d \in A_d : \|\zeta_d - \zeta^*_{n,d}\| \geq \epsilon} n^{-1} \sum_{s=1}^{n} \mathsf{E}_{\text{true}}[l_{s,d}(V_{s1}, \ldots, V_{s4}, \delta_{s1}, \ldots, \delta_{s4}; \zeta^*_{n,d}) - l_{s,d}(V_{s1}, \ldots, V_{s4}, \delta_{s1}, \ldots, \delta_{s4}; \zeta_d)] > 0.$$

(C3) The true (unknown) copula C has continuous partial derivatives.

(C4) (i) $l_{s,d}(u_1, \ldots, u_4, \delta_1, \ldots, \delta_4; \zeta_d)$ is a continuous function of $\zeta_d \in A_d$ for any $(u_1, \ldots, u_4) \in]0, 1[^4$.

(ii) Denote

$$L_s = \sup_{\zeta_d \in A_d} |l_{s,d}(V_{s1}, \ldots, V_{s4}, \delta_{s1}, \ldots, \delta_{s4}; \zeta_d)|$$

$$\nabla l_{s,d}(u_1, \ldots, u_4, \delta_1, \ldots, \delta_4; \zeta_d) = \frac{\partial l_{s,d}(u_1, \ldots, u_4, \delta_1, \ldots, \delta_4; \zeta_d)}{\partial \zeta_d}$$

$$L_{s\zeta_d} = \sup_{\zeta_d \in A_d} |\nabla l_{s,d}(V_{s1}, \ldots, V_{s4}, \delta_{s1}, \ldots, \delta_{s4}; \zeta_d)|.$$

Then, $\lim_{K\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{s=1}^{n} \mathsf{E}_{\text{true}}[L_s I(L_s \geq K) + L_{s\zeta_d} I(L_{s\zeta_d} \geq K)] = 0$.

(iii) For any $\eta > 0$ and any $\epsilon > 0$, there is $K > 0$ such that $|l_{s,d}(u_1, \ldots, u_4, \delta_1, \ldots, \delta_4; \zeta_d)| \leq K |l_{s,d}(u_1', \ldots, u_4', \delta_1, \ldots, \delta_4; \zeta_d)|$ for all $\zeta_d \in A_d$ and all $u_r \in [\eta, 1[$ such that $1 - u_r \geq \epsilon(1 - u_r')$ $(r = 1, \ldots, 4)$.

(C5) If T_{sr} $(s = 1, \ldots, n)$ are subject to non-trivial censoring (i.e., $C_{sr} \neq \infty$), then $\widehat{S}_r$ is truncated at the tail in the sense that for some τ_r, $\widehat{S}_r(t_r) = \widehat{S}_r(\tau_r)$ for all $t_r \geq \tau_r$ and $\liminf_{n\to\infty} n^{-1} \sum_{s=1}^{n} S_{G_{sr}}(\tau_r) S_r(\tau_r) > 0$.

Condition (C1) provides a very general censoring setting, e.g., the censoring times $(C_{s1}, \ldots, C_{s4})$ are allowed to be non-identically distributed and can be discrete or continuous. Condition (C2) ensures that the pseudo-true parameter value $\zeta^*_{n,d}$ is uniquely identifiable, while condition (C5) is imposed to handle the possible tail instability of the Kaplan-Meier estimator. Conditions (C3) and (C4) are technical conditions.

Proposition B.1.1. *(Chen et al., 2010) Under conditions (C1)–(C5), we have:*

(1) $\|\widehat{\zeta}_{n,d} - \zeta^*_{n,d}\| = o_P(1)$,

(2) $n^{-1}\log\widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) = n^{-1}\mathsf{E}_{\text{true}}[\log L_{n,d}(\zeta^*_{n,d})] + o_P(1)$.

Proposition B.1.1 (1) states that the maximum quasi-loglikelihood estimator $\widehat{\zeta}_{n,d}$ is a consistent estimator for the pseudo-true parameter value $\zeta^*_{n,d}$, while Proposition B.1.1 (2) is a weak law of large numbers.

Denote

$$\begin{aligned}\nabla^2 l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d) &= \frac{\partial^2 l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d)}{\partial\zeta_d^2}\\ \nabla_r l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d) &= \frac{\partial l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d)}{\partial\zeta_d\partial u_r}\end{aligned}$$

and

$$\begin{aligned}I_r(Y_{sr},\delta_{sr})(Y_{tr}) &= -S_r(Y_{tr})\left[\int_{-\infty}^{Y_{tr}}\frac{dN_{sr}(u)}{P_{n,r}(u)} - \int_{-\infty}^{Y_{tr}}\frac{I(Y_{sr}\geq u)d\Lambda_r u}{P_{n,r}(u)}\right]\\ W_r(Y_{sr},\delta_{sr};\zeta^*_{n,d}) &= \mathsf{E}_{\text{true}}[\nabla_r l_{s,d}(V_{s1},\ldots,V_{s4},\delta_1,\ldots,\delta_4;\zeta^*_{n,d})I_r(Y_{sr},\delta_{sr})(Y_{tr})|Y_{sr},\delta_{sr}]\end{aligned}$$

with $\Lambda_r(u) = -\log(S_r(u))$, $N_{sr}(u) = \delta_{sr}I(Y_{sr}\leq u)$, $dN_{sr}(u) = N_{sr}(u) - N_{sr}(u-)$ and $P_{n,r}(u) = n^{-1}\sum_{s=1}^n P(Y_{sr}\geq u)$. Let $\mathsf{Var}_{\text{true}}$ denote the variance with respect to the true (unknown) copula C.

The following (technical) conditions are sufficient to ensure the asymptotic normality of the maximum quasi-loglikelihood estimator $\widehat{\zeta}_{n,d}$ (Chen *et al.*, 2010).

(A1) (i) (C2) holds with $\zeta^*_{n,d}\in\text{int}(A^*_d)$ for all n, where A^*_d is a compact subset of A_d.

(ii) $B_{n,d} = -n^{-1}\sum_{s=1}^n \mathsf{E}_{\text{true}}[\nabla^2 l_{s,d}(V_{s1},\ldots,V_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta^*_{n,d})]$ has all its eigenvalues bounded below and above by some finite positive constants.

(iii) $\Sigma_{n,d} = n^{-1}\sum_{s=1}^n \mathsf{Var}_{\text{true}}[\nabla l_{s,d}(V_{s1},\ldots,V_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta^*_{n,d}) + \sum_{r=1}^4 W_r(Y_{sr},\delta_{sr};\zeta^*_{n,d})]$ has all its eigenvalues bounded below and above by some finite positive constants.

(iv) $\nabla l_{s,d}(V_{s1},\ldots,V_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta^*_{n,d}) + \sum_{r=1}^4 W_r(Y_{sr},\delta_{sr};\zeta^*_{n,d})$ satisfies Lindeberg condition $(s=1,\ldots,n)$.

(A2) Functions $\nabla^2 l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d)$ and $\nabla_r l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d)$ are well-defined and continuous in $(u_1,\ldots,u_4,\zeta_d)\in]0,1[^4\times A_d$ $(r=1,\ldots,4)$.

(A3) (i) $|\nabla l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta^*_{n,d})| \leq b\prod_{r=1}^4\{u_r(1-u_r)\}^{-a_r}$ for some $b>0$ and $a_r \geq 0$ such that $\limsup_{n\to\infty} n^{-1}\mathsf{E}_{\text{true}}[\prod_{r=1}^4\{V_{sr}(1-V_{sr})\}^{-2a_r}] < \infty$.

(ii) $|\nabla_r l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta^*_{n,d})| \leq p\prod_{r=1}^4\{u_r(1-u_r)\}^{-a_r}$ for some p and a_r such that $\limsup_{n\to\infty} n^{-1}\mathsf{E}_{\text{true}}[\{V_{sr}(1-V_{sr})\}^{\xi_r-a_r}\prod_{k\neq r=1}^4\{V_{sk}(1-V_{sk})\}^{a_k}] < \infty$ for some $\xi_r \in]0,1/2[$.

(A4) (i) Denote

$$\begin{aligned} L_{sr\zeta_d} &= \sup_{\zeta_d\in A_d}|\nabla_r l_{s,d}(V_{s1},\ldots,V_{s4},\delta_1,\ldots,\delta_4;\zeta_d)| \\ L_{s\zeta_d^2} &= \sup_{\zeta_d\in A_d}|\nabla^2 l_{s,d}(V_{s1},\ldots,V_{s4},\delta_1,\ldots,\delta_4;\zeta_d)|. \end{aligned}$$

Then, $\lim_{K\to\infty}\limsup_{n\to\infty} n^{-1}\sum_{s=1}^n \mathsf{E}_{\text{true}}[L_{sr\zeta_d}I(L_{sr\zeta_d} \geq K) + L_{s\zeta_d^2} I(L_{s\zeta_d^2}\geq K)] = 0$.

(ii) For any $\eta>0$ and any $\epsilon>0$, there is $K>0$ such that

$$\begin{aligned} &|\nabla l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d)| + |\nabla^2 l_{s,d}(u_1,\ldots,u_4,\delta_1,\ldots,\delta_4;\zeta_d)| \\ &\leq K\{|\nabla l_{s,d}(u_1',\ldots,u_4,\delta_1,\ldots,\delta_4';\zeta_d)| + |\nabla^2 l_{s,d}(u_1',\ldots,u_4',\delta_1,\ldots,\delta_4;\zeta_d)|\} \end{aligned}$$

for all $\zeta_d\in A_d$ and all $u_r \in [\eta,1[$ such that $1-u_r \geq \epsilon(1-u_r')$ $(r=1,\ldots,4)$.

Proposition B.1.2. *(Chen et al., 2010) Under conditions (C1)–(C5) and (A1)–(A4), we have:* $B_{n,d}\Sigma_{n,d}^{-1/2}\sqrt{n}(\widehat{\zeta}_{n,d}-\zeta^*_{n,d}) \to N(0,I_{q_d})$ *in distribution, where* $B_{n,d}$ *and* $\Sigma_{n,d}$ *are defined in (A1).*

Denote by e_{ds} the difference of the loglikelihood component evaluated at the pseudo-true parameter value and its expected value under the true (unknown) copula C ($d=1,\ldots,D$ and $s=1,\ldots,n$) and define for $d\neq d'=1,\ldots,D$:

$$\sigma_{d,d'} = n^{-1}\sum_{s=1}^n \mathsf{E}_{\text{true}}[(e_{ds}-\mathsf{E}[e_{ds}])(e_{d's}-\mathsf{E}[e_{d's}])].$$

Proposition B.1.3. *(Chen et al., 2010) Assume that conditions (C1)–(C5) and (A1)–(A4) are satisfied for all models* M_d *(*$d=1,\ldots,D$*) and that* e_{ds} *satisfies a Lindeberg condition. If* $\Omega_n = (\sigma_{d,d'})$ *(*$d\neq d'=1,\ldots,D$*) is finite and its largest eigenvalue is positive uniformly in* n*, then*

$$\begin{aligned} n^{-1/2}\Big[&\log\widetilde{L}_{n,d}(\widehat{\zeta}_{n,d}) - \log\widetilde{L}_{n,d'}(\widehat{\zeta}_{n,d'}) \\ &-\mathsf{E}_{\text{true}}[\log L_{n,d}(\zeta^*_{n,d}) - \log L_{n,d'}(\zeta^*_{n,d'})]\Big]_{d\neq d'=1,\ldots,D} \to N(0,\Omega_n) \end{aligned}$$

in distribution.

The extra Lindeberg condition on e_{ds} is needed to allow for non-nested models.

B.1.2 Theorems used in Section 5.7

The calculation of the tail dependence parameters in Section 5.7 is based on findings in Joe and Hu (1996) and Nelsen (2006). In this section we state the required theorems.

Theorem B.1.1. *(Nelsen, 2006) The tail behavior of* (5.12) *can be described by*

$$\delta_L = \lim_{s\to\infty} \frac{\widetilde{\psi}(2s)}{\widetilde{\psi}(s)} \text{ and } \delta_U = 2 - \lim_{s\to 0} \frac{1-\widetilde{\psi}(2s)}{1-\widetilde{\psi}(s)}.$$

Theorem B.1.2. *(Joe and Hu, 1996) The tail behavior of* (5.13) *can be described by*

(1) if the lower tail behavior of K_{ij} is given by $K_{ij}(u,u) \sim \beta u^{\rho}$ as $u \to 0$ ($\rho \geq 1$) then

$$\delta_L = \gamma \lim_{s\to\infty} \frac{\psi^{'}(-\log\beta + \gamma s)}{\psi^{'}(s)}$$

where $\gamma = 1/3(\rho+4) \geq 1$,

(2)

$$\delta_U = 2 - \gamma \lim_{s\to 0} \frac{\psi^{'}(\gamma s)}{\psi^{'}(s)}$$

where $\gamma = 2 - \beta/3 \in [1,2]$ and $\beta \in [0,1]$ denotes the upper tail dependence of K_{ij}.

Theorem B.1.3. *(Joe and Hu, 1996) The tail behavior of* (5.14) *can be described by*

(1) if the lower tail behavior of K_{ij} is given by $K_{ij}(u,u) \sim \beta u^{\rho}$ as $u \to 0$ ($\rho \geq 1$) then

$$\delta_L = \rho \lim_{s\to\infty} \frac{\psi^{'}(-\log\beta + \rho s)}{\psi^{'}(s)},$$

(2)

$$\delta_U = 2 - \gamma \lim_{s\to 0} \frac{\psi^{'}(\gamma s)}{\psi^{'}(s)}$$

where $\gamma = 2 - \beta \in [1,2]$ and $\beta \in [0,1]$ denotes the upper tail dependence of K_{ij}.

B.2 Lemmas used in Chapter 6

The proofs of the theorems in Section 6.4 and Section 6.5 rely on the asymptotic behavior of the Kaplan-Meier estimator and its quantiles. In this section we collect the required properties.

Consider univariate right-censored event time data. Denote the observed time of item r ($r = 1,\ldots,n$) by $Y_r = \min(T_r, C_r)$ with T_r the true event time and C_r the

censoring time. The indicator $\delta_r = I(T_r \leq C_r)$ equals one if $Y_r = T_r$ and zero otherwise. Event times and censoring times are assumed to be independent. Denote the distribution function of T (C) by F (G) and the survival function by S (S_G). With H the distribution function of Y, we obtain $1 - H = (1 - F)(1 - G)$. Further, for the upper endpoint of the support of F, G and H, we have $T_H = \min(T_F, T_G)$.

Recall, the Kaplan-Meier estimator (Kaplan and Meier, 1958) for S is given by

$$\widehat{S}(t) = \prod_{r:Y_{(r)} \leq t} \Big(\frac{n-r}{n-r+1}\Big)^{\delta_{(r)}}$$

where $Y_{(1)} \leq \ldots \leq Y_{(n)}$ are the order statistics of $Y_1, \ldots, Y_n$ and $\delta_{(1)}, \ldots, \delta_{(n)}$ are the corresponding δ_r's. The quantile functions of S and $\widehat{S}$ are defined in the usual way: for $0 < u < 1$, $S^{-1}(u) = \inf\{t : S(t) \leq u\}$ and $\widehat{S}^{-1}(u) = \inf\{t : \widehat{S}(t) \leq u\}$.

The first two lemmas are on the consistency of $\widehat{S}$ and $\widehat{S}^{-1}$. We need the conditions:

(C1) S and S_G are continuous.

(C9) For $S(T_G) < u < 1$, S is differentiable at $S^{-1}(u)$ with $f(S^{-1}(u)) > 0$ where $f = -S'$.

Note that the numbering of the conditions is the same as in Section 6.3. The formulation is however adapted to the present univariate situation.

Lemma B.2.1. *(Földes and Rejtő, 1981). Assume (C1). Then, for $T < T_H$,*

$$\sup_{0 \leq t \leq T} |\widehat{S}(t) - S(t)| = O(n^{-1/2}(\log\log n)^{1/2}) \text{ a.s.}$$

Lemma B.2.2. *(Gijbels, 1990). Assume (C1), (C9) and $S^{-1}(u) < T < T_H$. Let $\{a_n\}$ be a sequence of positive numbers tending to zero and, for n sufficiently large,*

$$a_n^2 n(\log n)^{-1} f^2(S^{-1}(u))(1 - H(T))^4/72 > 1.$$

Then,

$$|\widehat{S}^{-1}(u) - S^{-1}(u)| \leq a_n = O(n^{-1/2}(\log n)^{1/2}) \text{ a.s.}$$

The next two lemmas provide an i.i.d. representation for $\widehat{S}$ and $\widehat{S}^{-1}$.

Lemma B.2.3. *(Lo and Singh, 1986; Major and Retjő, 1988). Assume (C1). Then, for* $0 \le t \le T < T_H$,

$$\widehat{S}(t) = S(t) - \frac{1}{n}\sum_{r=1}^{n}\psi_r(t) + R_n(t)$$

with $\sup_{0\le t\le T}|R_n(t)| = O(n^{-1}\log n)$ *a.s. and*

$$\begin{aligned}\psi_r(t) &= S(t)\Big[\int_0^t \frac{I(Y_r \le y) - H(y)}{(1-H(y))^2}dH^u(y) \\ &+ \frac{I(Y_r \le t, \delta_r = 1) - H^u(t)}{1-H(t)} - \int_0^t \frac{I(Y_r \le y, \delta_r = 1) - H^u(y)}{(1-H(y))^2}dH(y)\Big]\end{aligned}$$

where $H^u(t) = P(Y \le t, \delta = 1)$.

Lemma B.2.4. *(Gijbels and Veraverbeke, 1998). Assume (C1) and (C9). Then, for* $S^{-1}(u) < T < T_H$,

$$\widehat{S}^{-1}(u) = S^{-1}(u) - \frac{1}{n}\frac{1}{f(S^{-1}(u))}\sum_{r=1}^{n}\psi_r(S^{-1}(u)) + o_P(n^{-1/2})$$

with ψ_r *as in Lemma B.2.3.*

The last two lemmas describe the oscillation behavior of $\widehat{S} - S$ and $S_{n,Y} - S_Y$ (defined in Section 6.4). The following conditions are needed:

(C2) S_Y is Lipschitz of order 1.

(C11) S is differentiable with $|S^{'}|$ bounded in $[0, T]$ with $T < T_H$.

Lemma B.2.5. *(Gijbels, 1990; Schäfer, 1986). Assume (C1) and (C11). Let* $\{a_n\}$ *be a sequence of positive numbers tending to zero with* $a_n n(\log n)^{-1} > c > 0$ *for* n *sufficiently large. Then,*

$$\sup_{0\le x,y\le T}\sup_{|x-y|\le a_n}|[\widehat{S}(x) - \widehat{S}(y)] - [S(x) - S(y)]| = O(a_n^{1/2}n^{-1/2}(\log n)^{1/2}) \text{ a.s.}$$

Lemma B.2.6. *Assume (C2). Let* $0 < u_1, u_2 < 1$ *and let* $\{a_n\}$ *be a sequence of positive numbers such that, for some constant* $c > 0$, $a_n \sim cn^{-1/2}(\log n)^{1/2}$ *as* $n \to \infty$. *Then,*

$$\begin{aligned}\sup_{|x|\le a_n, |y|\le a_n} &|[S_{n,Y}(S_1^{-1}(u_1) + x, S_2^{-1}(u_2) + y) - S_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2))] \\ &-[S_Y(S_1^{-1}(u_1) + x, S_2^{-1}(u_2) + y) - S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2))]| \\ &= O(n^{-3/4}(\log n)^{3/4}) \text{ a.s.} \qquad \text{(B.1)}\end{aligned}$$

Proof. Let $\{b_n\}$ be a sequence of positive numbers such that, for some constant $c_0 > 0$, $b_n \sim c_0 n^{1/4}(\log n)^{1/2}$ as $n \to \infty$. Define $\theta_{k,n} = S_1^{-1}(u_1) + a_n b_n^{-1} k$ and $\theta'_{l,n} = S_2^{-1}(u_2) + a_n b_n^{-1} l$ for $k, l = -b_n, \ldots, b_n$.

Due to the monotonicity of S_Y and $S_{n,Y}$ in both arguments, the left hand side of (B.1) is bounded above by $K_n + L_n$ with

$$K_n = \max_{-b_n \le k,l \le b_n} |[S_{n,Y}(\theta_{k,n}, \theta'_{l,n}) - S_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2))] \\ -[S_Y(\theta_{k,n}, \theta'_{l,n}) - S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2))]|$$

and

$$L_n = \max_{-b_n \le k,l \le b_n - 1} |S_Y(\theta_{k+1,n}, \theta'_{l+1,n}) - S_Y(\theta_{k,n}, \theta'_{l,n})|.$$

Consider L_n. Due to the Lipschitz continuity of S_Y, it holds that

$$\begin{aligned} |S_Y(\theta_{k+1,n}, \theta'_{l+1,n}) - S_Y(\theta_{k,n}, \theta'_{l,n})| &\le |\theta_{k+1,n} - \theta_{k,n}| + |\theta'_{l+1,n} - \theta'_{l,n}| \\ &= |a_n b_n^{-1}| + |a_n b_n^{-1}| = 2a_n b_n^{-1}. \end{aligned}$$

Therefore, $L_n = O(a_n b_n^{-1}) = O(n^{-3/4})$.

Consider K_n. We want to show that $K_n = O(n^{-3/4}(\log n)^{3/4})$ a.s. By the Borel-Cantelli lemma the latter holds if

$$\sum_{n=1}^{\infty} P(K_n \ge \gamma_n) < \infty$$

where $\gamma_n = c_1 n^{-3/4}(\log n)^{3/4}$ with $c_1 > 0$ a constant (to be specified later).

To prove the above, we rely on

$$P(K_n \ge \gamma_n) \le \sum_{k,l=-b_n}^{b_n} P(G_{kl,n} \ge \gamma_n)$$

where

$$G_{kl,n} = |[S_{n,Y}(\theta_{k,n}, \theta'_{l,n}) - S_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2))] \\ -[S_Y(\theta_{k,n}, \theta'_{l,n}) - S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2))]|.$$

Note that $nG_{kl,n} = |\sum_{s=1}^{n} Y_s - \sum_{s=1}^{n} E(Y_s)|$ with

$$Y_s = I(Y_{s1} > \theta_{k,n}, Y_{s2} > \theta'_{l,n}) - I(Y_{s1} > S_1^{-1}(u_1), Y_{s2} > S_2^{-1}(u_2)).$$

Indeed,

$$\begin{aligned}\sum_{s=1}^{n} Y_s &= \sum_{s=1}^{n}\Big[I(Y_{s1} > \theta_{k,n}, Y_{s2} > \theta_{l,n}^{'}) - I(Y_{s1} > S_1^{-1}(u_1), Y_{s2} > S_2^{-1}(u_2))\Big] \\ &= \sum_{s=1}^{n} I(Y_{s1} > \theta_{k,n}, Y_{s2} > \theta_{l,n}^{'}) - \sum_{s=1}^{n} I(Y_{s1} > S_1^{-1}(u_1), Y_{s2} > S_2^{-1}(u_2)) \\ &= nS_{n,Y}(\theta_{k,n}, \theta_{l,n}^{'}) - nS_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2))\end{aligned}$$

and

$$\begin{aligned}\sum_{s=1}^{n} E(Y_s) &= E\Big[nS_{n,Y}(\theta_{k,n}, \theta_{l,n}^{'}) - nS_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2))\Big] \\ &= E\Big[nS_{n,Y}(\theta_{k,n}, \theta_{l,n}^{'})\Big] - E\Big[nS_{n,Y}(S_1^{-1}(u_1), S_2^{-1}(u_2))\Big] \\ &= nS_Y(\theta_{k,n}, \theta_{l,n}^{'}) - nS_Y(S_1^{-1}(u_1), S_2^{-1}(u_2))\end{aligned}$$

where the last equality holds since $nS_{n,Y}(y_1, y_2)$ follows a Binomial distribution with parameters n and $S_Y(y_1, y_2)$.

We apply Bernstein's inequality to obtain

$$\begin{aligned}P(nG_{kl,n} \geq n\gamma_n) &= P(G_{kl,n} \geq \gamma_n) \\ &\leq 2\exp\left(-\frac{n^2\gamma_n^2}{2\sum_{s=1}^{n} \mathsf{Var}(Y_s) + 4n\gamma_n/3}\right).\end{aligned}$$

Further,

$$\begin{aligned}\mathsf{Var}(Y_s) &\leq E(Y_s^2) \\ &= E\Big[I(Y_{s1} > \theta_{k,n}, Y_{s2} > \theta_{l,n}^{'})\Big] + E\Big[I(Y_{s1} > S_1^{-1}(u_1), Y_{s2} > S_2^{-1}(u_2))\Big] \\ &\qquad -2E\Big[I(Y_{s1} > \theta_{k,n}, Y_{s2} > \theta_{l,n}^{'}, Y_{s1} > S_1^{-1}(u_1), Y_{s2} > S_2^{-1}(u_2))\Big] \\ &= S_Y(\theta_{k,n}, \theta_{l,n}^{'}) + S_Y(S_1^{-1}(u_1), S_2^{-1}(u_2)) \\ &\qquad -2S_Y(\min(\theta_{k,n}, S_1^{-1}(u_1)), \min(\theta_{l,n}^{'}, S_2^{-1}(u_2))) \\ &\leq c_2 a_n\end{aligned}$$

where the last inequality holds due to the Lipschitz continuity of S_Y.

Hence,

$$\begin{aligned}
& \frac{n^2\gamma_n^2}{2\sum_{s=1}^n \mathsf{Var}(Y_s) + 4n\gamma_n/3} \\
\geq\ & \frac{n^2\gamma_n^2}{2\sum_{s=1}^n c_2 a_n + 4n\gamma_n/3} \\
=\ & \frac{n\gamma_n^2}{2c_2 a_n + 4\gamma_n/3} \\
\geq\ & \frac{nc_1^2 n^{-3/2}(\log n)^{3/2}}{(2c_2 c + 4c_1/3)n^{-1/2}(\log n)^{1/2}} \qquad \text{for } n \text{ sufficiently large} \\
=\ & \frac{c_1^2}{2c_2 c + 4c_1/3}\log n.
\end{aligned}$$

Given c and c_2, we can choose c_1 sufficiently large such that $\frac{c_1^2}{2c_2 c + 4c_1/3} \geq 2$. We then have

$$\begin{aligned}
& \frac{n^2\gamma_n^2}{2\sum_{s=1}^n \mathsf{Var}(Y_s) + 4n\gamma_n/3} \geq 2\log n \\
\Rightarrow\ & -\frac{n^2\gamma_n^2}{2\sum_{s=1}^n \mathsf{Var}(Y_s) + 4n\gamma_n/3} \leq -2\log n \\
\Rightarrow\ & 2\exp\left(-\frac{n^2\gamma_n^2}{2\sum_{s=1}^n \mathsf{Var}(Y_s) + 4n\gamma_n/3}\right) \leq 2n^{-2}.
\end{aligned}$$

Consequently,

$$P(K_n \geq \gamma_n) \leq \sum_{k,l=-b_n}^{b_n} P(G_{kl,n} \geq \gamma_n) \leq \sum_{k,l=-b_n}^{b_n} 2n^{-2} = 8b_n^2 n^{-2}$$

and thus $P(K_n \geq \gamma_n) = O(b_n^2 n^{-2}) = O(n^{-3/2}\log n) = O(n^{-10/9})$. Moreover, $\sum_{n=1}^{\infty} P(K_n \geq \gamma_n) \leq \sum_{n=1}^{\infty} n^{-10/9} < \infty$ and thus $K_n = O(n^{-3/4}(\log n)^{3/4})$.

Combining the results on K_n and L_n completes the proof. □

Note that the above is a two-dimensional version of a result by Bahadur (1966) (see also Serfling (1980)).

Samenvatting

In de overlevingsanalyse is de tijd tot een bepaalde gebeurtenis, de overlevingstijd, de stochastische variabele waarin men geïnteresseerd is. Voorbeelden van overlevingstijden zijn: (i) de tijd tot de ontwikkeling van hepatitis, (ii) de tijd tot de terugkeer van borstkanker en (iii) de tijd tot de genezing van een uierinfectie bij koeien. Het toepassingsdomein van de overlevingsanalyse is zeer breed. Vaak is de overlevingstijd onderworpen aan rechtse censurering. De overlevingstijd is dan, voor een aantal observaties, niet exact gekend en er wordt enkel een ondergrens geobserveerd (bv., de studie stopt nog voor hepatitis zich ontwikkelt). Veelal komen overlevingstijden gegroepeerd voor. Voorbeelden van groepering zijn: (i) een persoon wordt opgevolgd tot de ontwikkeling van diverse vormen hepatitis (A/B/C), (ii) de terugkeer van borstkanker wordt nagegaan bij patiënten die behandeld zijn in eenzelfde ziekenhuis en (iii) de uier van een koe bestaat uit vier kwartieren, de infectiestatus van elk kwartier wordt opgevolgd. De groep of cluster is hierbij achtereenvolgens (i) de persoon, (ii) het ziekenhuis en (iii) de koe.

De overlevingstijden binnen een cluster zijn doorgaans gelijkaardig; ze vertonen een zekere correlatie. Klassieke methoden uit de overlevingsanalyse die de associatie gepast in rekening brengen zijn het frailty model en het copula model. Een frailty model is een uitbreiding van het uitvalsmodel (hazard model) van Cox, waarbij een multiplicatieve stochastische factor, de frailty, aan de uitvalsfunctie wordt toegevoegd. Een frailty model is dus een conditioneel model, waarbij de frailty de afhankelijkheid tussen de observaties binnen een cluster bepaalt. In een copula model wordt de gezamelijke overlevingsfunctie aan de marginale overlevingsfuncties gekoppeld via een copula functie. De copula beschrijft de associatie tussen de observaties binnen een cluster. Een copula model wordt gebruikt voor de analyse van overlevingstijden indien de clusters klein en van gelijke grootte zijn, terwijl een frailty model ook toegepast kan worden indien de clusters

omvangrijk en/of van verschillende grootte zijn.

Het doel van dit proefschrift is methoden te ontwikkelen die toelaten om de afhankelijkheidsstructuur in rechts gecensureerde gegroepeerde overlevingstijden flexibel te modelleren. Hiertoe bestuderen we het frailty model en het copula model in meer detail.

In Deel 1 bekijken we het gedeeld (shared) frailty model. Hoofdstuk 2 beschrijft beknopt het Cox model en de uitbreiding naar het gedeeld frailty model. In een gedeeld frailty model bepaalt de keuze van de frailty dichtheid het type afhankelijkheid tussen de overlevingstijden. Mogelijke frailty dichtheden zijn, onder meer, de invers Gaussische en de positief stabiele dichtheid, de meest populaire is echter de gamma dichtheid.

In Hoofdstuk 3 ontwikkelen we, voor een gedeeld frailty model met een parametrische uitvalsfunctie als basis, een omnibus toets om de hypothese van een gamma frailty dichtheid na te gaan. Hiertoe definiëren we, via een orthonormale reeksontwikkeling, een nieuwe klasse van veralgemeende gamma frailty dichtheden. De marginale aannemelijkheidsfunctie voor rechts gecensureerde overlevingstijden heeft een expliciete vorm. Een orde selectie toets met bijbehorend bootstrap algoritme laat toe om binnen een reeks van veralgemeende gamma frailty dichtheden de meest geschikte dichtheid te vinden. Via een simulatiestudie onderzoeken we het onderscheidingsvermogen van de toets. Uit de resultaten blijkt dat de voorgestelde procedure erin slaagt om afwijkingen van een gamma frailty dichtheid te detecteren.

In Deel 2 bestuderen we het copula model. Hoofdstuk 4 beschrijft kort de basis van de copula theorie. Het sleutelresultaat hierin is de stelling van Sklar (Sklar, 1959) die aangeeft dat de copula functie het type afhankelijkheid tussen de overlevingstijden bepaalt.

In Hoofdstuk 5 ligt de focus op semiparametrische copulas: de marginale overlevingsfuncties worden niet-parametrisch gemodelleerd, terwijl de copula een parametrische vorm aanneemt. We starten met een vergelijking tussen de vaak gebruikte (geneste) Archimedische copulas (bv., Clayton en Gumbel) en de minder gekende Joe-Hu copulas (Joe en Hu, 1996). We tonen hierbij aan dat de afhankelijkheidstructuur horende bij een (geneste) Archimedische copula uiterst beperkt is en dat een Joe-Hu copula de vereiste flexibiliteit toelaat. Gezien de veelheid aan mogelijke copula modellen, is de ontwikkeling van een selectie methode noodzakelijk. Hiertoe beschouwen we, voor rechts gecensureerde overlevingstijden, een criterium dat gebaseerd is op de gepenaliseerde aannemelijkheidsfunctie en we bewijzen dat, onder bepaalde voorwaarden, dit criterium

hetzij het beste model (één model bereikt de kleinste Kullback-Leibler waarde - zwakke consistentie) hetzij het beste model met het laagste aantal parameters (er zijn meerdere modellen die een gelijkaardige kleine Kullback-Leibler waarde hebben - consistentie) verkiest. Het verband met de welbekende AIC en BIC criteria komt hierbij aan bod. Om de diverse copula modellen te fitten, gebruiken we een generisch R-programma waarin de componenten van de aannemelijkheidsfunctie numerisch benaderd worden via eindige voorwaartse differenties. Een simulatiestudie toont de goede werking van het R-programma aan.

Indien de keuze van de parametrische bouwstenen van een semiparametrische copula niet evident is, dan biedt het gebruik van een niet-parametrische copula een uitweg. Dit is het onderwerp van Hoofdstuk 6. We definiëren een nieuwe niet-parametrische copula schatter voor rechts gecensureerde overlevingstijden. Hierbij bekijken we twee vormen van censurering: (i) univariate censurering - per cluster is er één censureringstijd die gemeenschappelijk is voor alle componenten binnen een cluster en (ii) copula censurering - iedere component binnen een cluster heeft zijn eigen censureringstijd en de afhankelijkheid tussen de censureringstijden binnen een cluster wordt beschreven door een (gekende) parametrische copula. De situatie waarbij de censureringstijden binnen een cluster onafhankelijk zijn, is hierin bevat. Voor de nieuwe copula schatter tonen we de consistentie aan en geven we de asymptotische representatie als een som van onafhankelijke identisch verdeelde stochastische veranderlijken. Om de kwaliteit van de nieuwe copula schatter te evalueren, vergelijken we deze in diverse data scenario's met de recent ontwikkelde niet-parametrische copula schatter van Gribkova en Lopez (2014). De resultaten geven aan dat de hoeveelheid censurering, de sterkte van de associatie tussen de overlevingstijden alsook de grootte van de cluster bepalend zijn voor de kwaliteit van de schatter. Gegeven een concrete situatie, geniet de nieuwe copula schatter dan wel de schatter van Gribkova en Lopez (2014) de voorkeur.