University of Helsinki

Department of Mathematics and Statistics

Master's thesis

Estimation of the Tail Dependence Coefficient

Pekka Sohkanen

Autumn 2021

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty		Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science		Master's Program in Mathematics and Statistics

Tekijä — Författare — Author

Pekka Sohkanen

Työn nimi — Arbetets titel — Title

Estimation of the tail dependence coefficient

Oppiaine — Läroämne — Subject

Applied mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's Thesis	September 2021	82 s.

Tiivistelmä — Referat — Abstract

The fields of insurance and financial mathematics require increasingly intricate descriptors of dependency. In the realm of financial mathematics, this demand arises from globalisation effects over the past decade, which have caused financial asset returns to exhibit increasingly intricate dependencies between each other. Of particular interest are measurements describing the probabilities of simultaneous occurrences between unusually negative stock returns. In insurance mathematics, the ability to evaluate probabilities associated with the simultaneous occurrence of unusually large claim amounts can be crucial for both the solvency and the competitiveness of an insurance company. These sorts of dependencies are referred to by the term tail dependence.

In this thesis, we introduce the concept of tail dependence and the tail dependence coefficient, a tool for determining the amount of tail dependence between random variables. We also present statistical estimators for the tail dependence coefficient. Favourable properties of these estimators are investigated and a simulation study is executed in order to evaluate and compare estimator performance under a variety of distributions.

Some necessary stochastics concepts are presented. Mathematical models of dependence are introduced. Elementary notions of extreme value theory and empirical processes are touched on. These motivate the presented estimators and facilitate the proofs of their favourable properties.

Avainsanat — Nyckelord — Keywords

Tail dependence, Tail dependence estimation, Tail dependence coefficient

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpula Campus Library

Muita tietoja — Övriga uppgifter — Additional information

Contents

1 Introduction

Dependence modelling is a central concept in risk analysis. For example, insurance claims can be very dependent on each other, and in order to access related future risk, this dependence structure needs to be built into the mathematical model used for the risk analysis.

Independence between random variables has a rigorous mathematical definition, but dependence is a more intricate notion in that, barring the absence of independence, no one canonical mathematical measure of dependence exists. Instead, the existence and level of dependence between random variables can be described in a variety of ways.

Some forms of dependence may be more interesting than others for the particular purposes of the investigator. As an example of this, the dependence between extremal values of random variables, or "tail dependence", is a subject of study for which there is commercial interest. Insurance companies, for one, are particularly interested in the dependence between large claim amounts.

In this thesis, we introduce the tail dependence coefficients as tools for determining the amount of tail dependence between random variables. We present examples in which the tail dependence coefficients are calculated analytically for a variety of different joint distributions. We present multiple alternative representations for the tail dependence coefficients. We then use these different representations to motivate estimators for the tail dependence coefficients. We compare these estimators to each other in a simulation study.

Section 2 relates the mathematical basics needed for the following sections. Section 3 introduces the central concepts studied in this thesis and some mathematical results which connect these concepts with each other. Statistical estimators for these concepts are established in Section 4. The advantageous properties of these estimators are related and, for one specific estimator, proven. Section 5 recounts a simulation study into investigating the performance of these estimators in practice. Section 6 concludes with a brief overview of the findings of the simulation study. Appendix A is a glossary of notation. Appendix B contains the R-code which is used in the simulation study of Section 5.

2 Preliminary mathematics

2.1 Distribution functions and their generalised inverses

The random variables considered in this thesis are real-valued. Distribution functions of such random variables are non-decreasing and right-continuous. An inverse of the distribution function exists if the distribution function is continuous and strictly increasing. On the set of real numbers this corresponds to distributions whose support is an interval (strict monotony) and have no atoms (no jump-discontinuities).

Furthermore, a function $F\colon \mathbb{R} \to [0,1]$ can not be bijective, since it would necessarily be strictly monotonic and continuous, thus mapping $\mathbb{R}$ to an open interval, which its codomain, the closed interval $[0,1]$, is clearly not. Due to this, there does not exist an inverse to any distribution function in the strict sense of the definition of an inverse. To circumvent this problem, the codomain of F can simply be restricted to the open interval $(0,1)$ and inverses considered for this restriction.

However, it is possible to define inverses to distribution functions even when the support of the distribution is not the whole real line. In this case, the domain of the distribution function is restricted to the relevant interval. Denote

$$a_F := \inf\{x \in \mathbb{R} \mid F(x) > 0\} \quad \text{and} \quad b_F := \sup\{x \in \mathbb{R} \mid F(x) < 1\}.$$

Now an inverse distribution function is not strictly speaking an inverse for the function $F\colon \mathbb{R} \to [0,1]$, but rather a function $F^{-1}\colon (0,1) \to (a_F, b_F)$, which fulfils

$$\begin{aligned} F^{-1}(F(x)) &= x, \quad \text{for any } x \in (a_F, b_F) \quad \text{and} \\ F(F^{-1}(u)) &= u, \quad \text{for any } u \in (0,1). \end{aligned} \tag{2.1}$$

In the following, we generalise the concept of an inverse distribution function to have an analogue to the idea even in cases where no true inverse (in the sense of (2.1)) exists. The goal is also to generalise the above thinking so that we do not need to consider the specifics of the distribution e.g. on which interval F is injective and on which interval the inverse is defined.

Definition 2.1.1 Let $F\colon \mathbb{R} \to [0,1]$ be a distribution function. Its *generalised inverse* is the function

$$F^{\leftarrow} : \mathbb{R} \to \mathbb{R}, \quad F^{\leftarrow}(u) = \inf\{t \in \mathbb{R} \mid F(t) \geq u\}.$$

We will now investigate some properties of the generalised inverse.

Lemma 2.1.2 *Let X be a random variable with the distribution function F. Fix $u \in (0,1)$ and $x \in \mathbb{R}$.*

(i) $F(F^{\leftarrow}(u)) \geq u$. *Particularly*

$$F(F^{\leftarrow}(u)) = u \quad \Leftrightarrow \quad u \in \mathit{Range}(F).$$

(ii) $F^{\leftarrow}(F(x)) \leq x$. *Particularly*

$$F^{\leftarrow}(F(x)) = x \quad \Leftrightarrow \quad F(y) < F(x), \text{ for every } y < x.$$

(iii) $F(x) \geq u \quad \Leftrightarrow \quad x \geq F^{\leftarrow}(u)$

(iv) $F(x) < u \quad \Leftrightarrow \quad x < F^{\leftarrow}(u)$

(v) $F^{\leftarrow}$ *is non-decreasing.*

(vi) F *is continuous.* $\quad \Leftrightarrow \quad F^{\leftarrow}$ *is strictly increasing.*

Proof. (i): If $F(F^{\leftarrow}(u)) = u$, then trivially u is in the range of F. If $u \in (0,1)$ is in the range of F, there exists a real number $x \in \mathbb{R}$, such that,

$$F(x) = u \qquad \Longrightarrow \qquad x \in \{t \in \mathbb{R} \mid F(t) = u\}.$$

Now because F is increasing,

$$F^{\leftarrow}(u) = \inf\{t \in \mathbb{R} \mid F(t) \geq u\} = \inf\{t \in \mathbb{R} \mid F(t) = u\}.$$

That is, $F(F^{\leftarrow}(u)) = F(x) = u$. So the equality holds if and only if u is in the range of F

Suppose then that $u \in (0,1)$ is not in the range of F. Then

$$\{t \in \mathbb{R} \mid F(t) = u\} = \emptyset,$$

so

$$\{t \in \mathbb{R} \mid F(t) \geq u\} = \{t \in \mathbb{R} \mid F(t) > u\}.$$

We obtain the estimate

$$F(F^{\leftarrow}(u)) = F(\inf\{t \in \mathbb{R} \mid F(t) > u\}) > u.$$

Thus, we have shown claim (i).

(ii): Let $x \in \mathbb{R}$. Clearly

$$x \in \{t \in \mathbb{R} \mid F(t) \leq F(x)\}.$$

Since x belongs to the set, any lower bound of the set is going to be lesser than or equal to x. Therefore this order also applies to the largest lower bound, i.e. the infimum. Thus,

$$F^{\leftarrow}(F(x)) = \inf\{t \in \mathbb{R} \mid F(t) \geq F(x)\} \leq x.$$

Let us then investigate when the inequality is strict. Suppose $F^{\leftarrow}(F(x)) < x$. This happens exactly when $y := F^{\leftarrow}(F(x))$ is a number, for which it applies that $y < x$,

$$F(y) = F(F^{\leftarrow}(F(x))) \overset{\text{(i)}}{\geq} F(x),$$

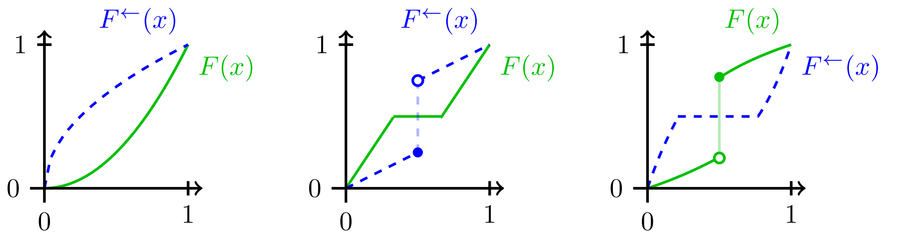
and since F is non-decreasing, the above means that $F(y) = F(x)$. The case where equality applies is the complement of this. That is, the equality applies when $F(y) < F(x)$ for all $y < x$.

(iii): Let $x \in \mathbb{R}$ and $u \in (0,1)$ so that $F(x) \geq u$. Then clearly

$$x \in \{t \in \mathbb{R} \mid F(t) \geq u\} \qquad \Longrightarrow \qquad x \geq \inf\{t \in \mathbb{R} \mid F(t) \geq u\} = F^{\leftarrow}(u).$$

Conversely suppose $x \in \mathbb{R}$ and $u \in [0,1]$ so that

$$x \geq F^{\leftarrow}(u) \qquad \Longrightarrow \qquad F(x) \geq F(F^{\leftarrow}(u)) \overset{\text{(i)}}{\geq} u.$$



(a) Bijective F.

(b) F stays constant on an interval.

(c) F has a jump discontinuity.

Figure 2.1: Example cases of different distribution functions F (drawn in green). As seen in (a), when F is bijective, the generalised inverse $F^{\leftarrow}$ (drawn in blue) coincides with the true inverse function. As illustrated in (b), intervals on which F stays constant translate to jump discontinuities in $F^{\leftarrow}$. Conversely, jump discontinuities in F lead to constancy intervals in $F^{\leftarrow}$. This is illustrated in (c).

(iv): Let $x \in \mathbb{R}$ and $u \in (0, 1)$. Then

$$F(x) < u \quad \overset{\text{(iii)}}{\iff} \quad x < F^{\leftarrow}(u).$$

(v): Let $0 < u < v < 1$. Then, by (i)

$$F(F^{\leftarrow}(v)) \geq v > u.$$

And so, by (iii)

$$F^{\leftarrow}(v) \geq F^{\leftarrow}(u).$$

This shows (v). (vi): Suppose F is continuous. Then, since it's a distribution function, it achieves all values in $(0, 1)$. Let $0 < u < v < 1$. Now u and v are both in the range of F. Thus, by (i),

$$F^{\leftarrow}(F(u)) = u < v = F^{\leftarrow}(F(v)).$$

So $F^{\leftarrow}$ is strictly increasing.

Suppose then that F is not continuous. In this case, according to (i), there exists a $u \in (0, 1)$, such that

$$v := F(F^{\leftarrow}(u)) > u.$$

By (v), $F^{\leftarrow}$ is non-decreasing. Therefore, the above implies

$$F^{\leftarrow}(u) \leq F^{\leftarrow}(v) = F^{\leftarrow}(F(F^{\leftarrow}(u))) \overset{\text{(ii)}}{\leq} F^{\leftarrow}(u).$$

This means that $F^{\leftarrow}$ is not strictly increasing. We have thus shown that the non-continuity of F implies that $F^{\leftarrow}$ is not strictly increasing. Consequently, we have shown the contraposition of this, which was our claim. □

From parts (i) and (ii) of Lemma 2.1.2, we deduce that when an inverse function exists in the sense of (2.1), it coincides with the general inverse.

Example 2.1.3 Suppose the restricted distribution function $F\colon \mathbb{R} \to [0, 1]$ has an inverse in the sense of (2.1). Denote this inverse by $F^{-1}\colon (0, 1) \to (a_F, b_F)$. With this notation, F is bijective

$(a_F, b_F) \to (0,1)$. So F attains all values in the interval $(0,1)$ and therefore, by part (i) of Lemma 2.1.2

$$F(F^{\leftarrow}(x)) = x, \quad \text{for all } x \in (0,1).$$

F is also an injective function $(a_F, b_F) \to (0,1)$. This means it is necessarily strictly monotonous. Since it is a distribution function, it is also non-decreasing. Together this means that F is strictly increasing. So by part (ii) of the same lemma

$$F^{\leftarrow}(F(x)) = x \quad \text{for all } x \in \mathbb{R}.$$

And thus for any $x \in (0,1)$

$$F^{\leftarrow}(x) = F^{-1}(x).$$

This example illustrates that the general inverse is convenient in the sense that removes the need to consider whether an inverse in the regular sense exists at all. Even when it does, the generalised inverse agrees with it on the relevant interval $(0,1)$ and is defined on the whole of $\mathbb{R}$.

We note some more general properties of distribution functions we'll use later.

Lemma 2.1.4 *Let X be a real random variable with a continuous distribution function F. For any $x \in \mathbb{R}$ we have*

$$\mathbb{P}(X \leq x) = \mathbb{P}(F(X) \leq F(x)).$$

Further, for random variables $X_i,\ i = 1, \ldots, d$ with continuous distribution functions $F_{X_i},\ i = 1, \ldots, d$, for any $x_1, \ldots, x_d \in \mathbb{R}$, it applies that

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = \mathbb{P}(F_{X_1}(X_1) \leq F_{X_1}(x_1)\ , \ldots,\ F_{X_d}(X_d) \leq F_{X_d}(x_d)).$$

Proof. For the first claim, let $x \in \mathbb{R}$ and denote

$$y := \sup\{t \in R \mid F(t) \leq F(x)\}.$$

According to part (ii) of Lemma 2.1.2, we have the inequality

$$F^{\leftarrow}(F(x)) \leq x \leq y,$$

of numbers in the domain of F. Since F is assumed to be continuous, we have $F(y) = F(x)$. By (i) of Lemma 2.1.2, we have $FF^{\leftarrow}F(x) = (FF^{\leftarrow})(F(x)) = F(x)$, since clearly, $F(x)$ is in the range of F. It follows then from the definitions of y and $F^{\leftarrow}$ that $[F^{\leftarrow}(F(x)), y]$ is precisely the interval in the domain of F which is mapped to $F(x)$. Thus

$$\{F(X) \leq F(x)\} = \{X \leq y\} = \{X \leq x\} \sqcup \{x < X \leq y\}.$$

And further

$$\begin{aligned}
\mathbb{P}(F(X) \leq F(x)) &= \mathbb{P}(X \leq x) + \mathbb{P}(x < X \leq y) \\
&= \mathbb{P}(X \leq x) + \mathbb{P}(x \leq X \leq y) \qquad (F \text{ is continuous.}) \\
&= \mathbb{P}(X \leq x) + F(y) - F(x) \\
&= \mathbb{P}(X \leq x).
\end{aligned}$$

This proves the first claim.

For the second claim, notice that according to part (vi) of Lemma 2.1.2, F_{X_i} is now strictly increasing for all $i = 1, \dots, d$. Thus

$$X_i \leq x_i \implies F_{X_i}(X_i) \leq F_{X_i}(x_i) \implies \{X_i \leq x_i\} \subset \{F_{X_i}(X_i) \leq F_{X_i}(x_i)\},$$

for every $i = 1, \dots, d$. And so also

$$\begin{aligned}
&\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \\
&\quad = (\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \cap \{X_i \leq x_i\}) \sqcup (\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \setminus \{X_i \leq x_i\}) \\
&\quad = \{X_i \leq x_i\} \sqcup (\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \setminus \{X_i \leq x_i\})
\end{aligned}$$

Since the union is disjoint, we can utilise the additivity of the probability measure and get

$$\mathbb{P}(\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \setminus \{X_i \leq x_i\}) = \mathbb{P}(F_{X_i}(X_i) \leq F_{X_i}(x_i)) - \mathbb{P}(X_i \leq x_i) = 0,$$

by the first part of this proof. So now

$$\begin{aligned}
\mathbb{P}(F_{X_1}(X_1) \leq F_{X_1}(x_1)\ , \dots,\ F_{X_d}(X_d) \leq F_{X_d}(x_d)) &= \mathbb{P}\left(\bigcap_{i=1}^{d} \{F_{X_i}(X_i) \leq F_{X_i}(x_i)\}\right) \\
= \mathbb{P}\left(\bigcap_{i=1}^{d} \left(\{X_i \leq x_i\} \sqcup (\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \setminus \{X_i \leq x_i\})\right)\right) \\
= \mathbb{P}\left(\bigcap_{i=1}^{d} \{X_i \leq x_i\}\right) \\
+ 2\sum_{i=1}^{d} \mathbb{P}\left(\{X_i \leq x_i\} \cap \left(\bigcap_{j \neq i}^{d} (\{F_{X_j}(X_j) \leq F_{X_j}(x_j)\} \setminus \{X_j \leq x_j\})\right)\right) \\
+ \mathbb{P}\left(\bigcap_{i=1}^{d} (\{F_{X_i}(X_i) \leq F_{X_i}(x_i)\} \setminus \{X_i \leq x_i\})\right).
\end{aligned}$$

The last two terms contain sets of measure zero in the intersection taken inside of the probability. These terms are thus equal to zero and we end up having shown the claim:

$$\mathbb{P}(F_{X_1}(X_1) \leq F_{X_1}(x_1)\ , \dots,\ F_{X_d}(X_d) \leq F_{X_d}(x_d)) = \mathbb{P}\left(\bigcap_{i=1}^{d} \{X_i \leq x_i\}\right)$$

□

Lemma 2.1.5 *Let X be a real-valued random variable with distribution function F. If F is continuous, then $F(X) \sim \mathcal{U}(0, 1)$.*

Proof. For continuous and strictly increasing F this result is elementary since in such a case there exists an inverse F^{-1}, at least in the sense of (2.1). Thus for any $u \in (0, 1)$,

$$\begin{aligned}
\mathbb{P}(F(X) \leq u) &= \mathbb{P}(F^{-1}(F(X)) \leq F^{-1}(u)) = \mathbb{P}(X \leq F^{-1}(u)) \\
&= F(F^{-1}(u)) = u.
\end{aligned}$$

Therefore we have shown $F(X)$ to have the distribution function of $\mathcal{U}(0,1)$.

Suppose F is only continuous and not necessarily strictly increasing. Let $u \in (0,1)$. Since F is a continuous distribution function, it achieves all values in $(0,1)$. We may therefore fix x such that $F(x) = u$ and notice

$$\{F(X) \leq u\} = \{X \leq F^{\leftarrow}(u)\} \cup \{X \in (F^{\leftarrow}(u), y]\}, \tag{2.2}$$

where $y = \sup\{z \in \mathbb{R} \mid F(z) \leq u\}$. Since F is continuous, $F(y) = F(x)$. Then, because $F(x)$ is in the range of F, by (i) of Lemma 2.1.2,

$$\begin{aligned}\mathbb{P}(X \in (F^{\leftarrow}(u), y]) &= F(y) - F(F^{\leftarrow}(u)) = F(y) - F(F^{\leftarrow}(F(x))) \\ &\overset{2.1.2\text{ (i)}}{=} F(y) - F(x) = 0.\end{aligned}$$

Now, because the sets in (2.2) are disjoint, we have the equality

$$\begin{aligned}\mathbb{P}(F(X) \leq u) &= \mathbb{P}(X \leq F^{\leftarrow}(u)) + \mathbb{P}(X \in (F^{\leftarrow}(u), y]) \\ &= F(F^{\leftarrow}(u)) + 0 \overset{2.1.2\text{ (i)}}{=} u,\end{aligned}$$

since u is in the range of F. This shows $F(X)$ has the distribution function of $\mathcal{U}(0,1)$. □

2.2 A result from extreme value theory

In extreme value theory, one is interested in questions concerning the maxima of random variables. Let $\{X_j\}_{j\in N}$ be a sequence of i.i.d. random variables with a common distribution function $F(x) = \mathbb{P}(X_j \leq x)$, for every $j \in \mathbb{N}$ and $x \in \mathbb{R}$. Denote

$$M_n := \max\{X_j \mid 1 \leq j \leq n\}.$$

Independence implies

$$F_{M_n}(x) = \mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) \overset{\perp\!\!\!\perp}{=} \prod_{j=1}^{n} \mathbb{P}(X_j \leq x) = (F(x))^n. \tag{2.3}$$

It follows from this, that the sequence of random variables M_n converge almost surely toward a constant. This can be seen by fixing $x < b_F = \sup\{t \in \mathbb{R} \mid F(t) < 1\}$ and noticing that clearly $F(x) < 1$. Thus

$$\lim_{n\to\infty} F_{M_n}(x) = \lim_{n\to\infty} (F(x))^n = 0. \tag{2.4}$$

It therefore turns out that

$$F_{M_n}(x) = \begin{cases} 0, & x < b_F \\ 1, & x \geq b_F, \end{cases}$$

that is,

$$\lim_{n\to\infty} M_n = b_F, \tag{2.5}$$

almost surely. It may be interesting to note that $b_F = ||X_1||_\infty$. Put simply, this means that the maximum M_n converges toward the largest value which X_1 attains with strictly positive probability. Though b_F is constant, it lies in infinity in many cases.

The convergence toward a constant expressed in (2.5) can be avoided by normalising the components of the sequence M_n. Regarding this, we get a central theorem in the field of extreme value theory

Theorem 2.2.1 *Let M_n be as above. Suppose there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ for all $n \in \mathbb{N}$, such that,*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \underset{n\to\infty}{\longrightarrow} G(x),$$

for all $x \in \mathbb{R}$, where G is a non-degenerate distribution function. Then G is the distribution function of one of three distributions:

(i) The Fréchet distribution: $\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases}$

for some $\alpha > 0$.

(ii) The Weibull distribution: $\Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x < 0 \\ 1, & x \geq 0 \end{cases}$

for some $\alpha > 0$.

(iii) The Gumbel distribution: $\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}$.

Proof. This theorem is proven, for example, in [16] (Proposition 0.3).

The distributions of Theorem 2.2.1 are referred to as *extreme value distributions*. If the sequences a_n and b_n fulfilling the conditions of this theorem exist, the underlying distribution function F is said to be in the *domain of attraction* of the extreme value distribution G.

Example 2.2.2 Let $\{X_j\}_{j\in\mathbb{N}}$, be a sequence of i.i.d. random variables so that $-X_j \sim \text{Exp}(1)$, for all $j = 1, \ldots, n$. Then for $x \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(M_n \leq x) &= \mathbb{P}(\max\{X_j \mid 1 \leq j \leq n\} \leq x) = \mathbb{P}(X_1 \leq x, \ \ldots, \ X_n \leq x) \\ &\overset{\perp\!\!\!\perp}{=} \prod_{j=1}^{n} \mathbb{P}(X_j \leq x) \overset{\text{i.i.d.}}{=} \big(\mathbb{P}(X_1 \leq x)\big)^n = \big(\mathbb{P}(-X_1 > -x)\big)^n. \end{aligned} \tag{2.6}$$

From here on we can simply deduce the distribution function of M_n by applying our knowledge of the distribution of $-X_1$. Suppose $x \geq 0$. Then $-x \leq 0$ and thus (2.6) becomes

$$\big(\mathbb{P}(-X_1 > -x)\big)^n = \big(1 - \mathbb{P}(-X_1 \leq -x)\big)^n = \big(1 - 0\big)^n = 1. \tag{2.7}$$

Suppose then that $x < 0$. Then $-x > 0$ and (2.6) gives

$$\big(\mathbb{P}(-X_1 > -x)\big)^n = e^{-(-x)n}. \tag{2.8}$$

Combining (2.7) and (2.8) gives us the distribution function of M_n over the whole real line.

$$\mathbb{P}(M_n \leq x) = \begin{cases} e^{-(-x)n}, & x < 0 \\ 1, & x \geq 0. \end{cases}$$

Like we observed in (2.4), the distribution function of M_n converges toward zero as n grows. In this case an obvious choice of normalising sequences is $a_n = 1/n$ and $b_n = 0$, for all $n \in \mathbb{N}$, since then

$$\mathbb{P}\left(\frac{M_n - 0}{1/n} \leq x\right) = \mathbb{P}\left(M_n \leq \frac{x}{n}\right) = \begin{cases} e^{-(-x)}, & x < 0 \\ 1, & x \geq 0, \end{cases} \tag{2.9}$$

for every $n \in \mathbb{N}$. This is the Weibull distribution with $\alpha = 1$.

Theorem 2.2.1 guarantees that if suitable sequences a_n and b_n exist, the resulting distribution is an extreme value distribution. In Example 2.2.2, these normalising sequences are easily determined. It is important to note, however, that this example is academic and finding such sequences for any given M_n is rarely trivial.

2.3 Empirical distribution functions

In statistics, random variables are used to model indeterminate phenomena. They describe a scenario where some process leads to the realisation of a value. The inner workings of this process are not known well enough for the resulting value to be predictable, so we model the resulting value as being random. This is to say, it is characterised by the probabilities associated with the realisations of its possible values.

A probabilistic model is fitted into data emergent from the phenomenon under investigation, and the model is then used to predict future resulting values of the phenomenon. The amount of data is, of course, always going to be finite. Thus models are fit into data using a variety of statistical methods. The distribution underlying a sample i.i.d. of observations can also be estimated with what is called an empirical distribution.

Definition 2.3.1 Let $X_i, i = 1, \ldots, n$ be i.i.d. random variables. Here, they denote a sample of observations of their common distribution. Set

$$\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}.$$

Then $\hat{F}_n$ is called the *empirical distribution function* constructed from the observations $X_i, i = 1, \ldots, n$. The subindex n denotes the size of the sample used to construct the empirical distribution function.

Denote the common distribution function of the observations as F. Proving point-wise convergence of the empirical distribution function to the common distribution function is rather straightforward. For any $x \in \mathbb{R}$,

$$\mathbb{E}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(X_i \leq x) = \mathbb{P}(X_1 \leq x) = F(x).$$

And particularly because this expected value exists, the strong law of large numbers states that for any $x \in \mathbb{R}$ the convergence result

$$\hat{F}_n(x) \xrightarrow{\text{a.s.}} F(x),$$

holds. Thus we have almost sure convergence. However, there is a stronger result stating that this convergence is, in fact, uniform. This result is sometimes referred to as the Glivenko-Cantelli

Theorem.

Theorem 2.3.2 *Let $X_i, i = 1, \dots, n$ be i.i.d. random variables with common distribution function F and empirical distribution function $\hat{F}_n$. Then*

$$\sup_{x \in \mathbb{R}} \left| F(x) - \hat{F}_n(x) \right| \xrightarrow[n \to \infty]{a.s.} 0.$$

Proof. The proof can be found, for example, in [19] (Theorem 19.1).

3 Dependence structures in risk analysis

According to the mathematical definition of independence, random variables are either independent or they are not. The absence of independence can be seen as an indication of the presence of dependence, but the mere knowledge of the lack of independence is hardly very useful for risk analysis purposes. Random variables could be considered dependent on each other or not, but a question worth considering is, how can one quantify the level of dependence between them? There are various ways of measuring dependence in this quantitative sense. A well-known example of such a measurement is the concept of covariance.

Covariance is relatively easy to calculate, even from large data sets, and it gives a single value describing the dependence relationship under inquiry. This being said, considered as a measurement of dependence, covariance has its limitations. When introducing covariance as a concept in a probability or statistics course, a classical remark to make is that completely uncorrelated random variables can still be dependent on each other (the classical example being the random pair (X, Y), where $X \sim \mathcal{N}(0,1)$ and $Y = X^2$).

This does not, of course, make covariance a useless tool, but it tells one that in simplifying the dependence relationship to one number, at least some nuance is lost. Next, we observe the insufficiency of covariance in describing tail dependencies, that is, as a measurement of the likelihood associated with simultaneous occurrence of values located at the tails of distributions. In order to do this, we must first introduce some other way of speaking about such dependence relationships. One such concept is the tail dependence coefficient.

3.1 Tail dependence

Let X and Y be real-valued random variables. When asking questions about the simultaneous occurrence of their unlikely values, it is natural to consider something such as

$$\mathbb{P}(|X| > u \ , \ |Y| > u), \tag{3.1}$$

for large values of u. The value (3.1) describes how much probability mass in the joint distribution of X and Y is allotted to the situation where they are simultaneously further than u away from the origin.

Investigation into tail dependence is largely motivated by its applications into e.g. finance and actuarial mathematics. In both of these cases, one is most interested in the simultaneous occurrence of either 'very large' values or 'very small' ones, whatever the specifics defining these qualitative terms may be. So in the consideration of tail dependence in this work, the absolute values in (3.1) are dropped and questions concerning the probability of events such as

$$\{X > u\} \cap \{Y < -u\} \quad \text{or} \quad \{X < -u\} \cap \{Y > u\}$$

are not considered. Instead, we are interested in the positive values of X and Y which are close to, or far away from, 0. In this thesis, we concentrate on events concerning the upper tail, and so the events considered are mostly of the form

$$\{X > u\} \cap \{Y > u\}.$$

Further, we can not necessarily always know which values of u make (3.1) interesting. For large enough u the value (3.1) becomes close to zero, irrespective of the distributions of X and Y. To get rid of this problem we instead consider the limit of a conditional probability. In this spirit, we consider values of the form

$$\lim_{u\to\infty} \mathbb{P}(X > u \mid Y > u) = \lim_{u\to\infty} \frac{\mathbb{P}(X > u \ , \ Y > u)}{\mathbb{P}(Y > u)}. \tag{3.2}$$

Here it is assumed that $\mathbb{P}(Y > u) > 0$, for arbitrarily large u, otherwise the conditional expectation in the left hand side of 3.2 would be set to 0. The limit 3.2 tells us something about the tail behaviour of X related to the tail behaviour of Y. It is not necessarily equal to zero even though u tends to infinity, but can achieve any value in the interval $[0, 1]$. As such, (3.2) is a much more fitting candidate for a measurement of tail dependence than (3.1) is.

The limit (3.2) still exhibits unwanted behaviour, in particular it is asymmetric when applied to certain random pairs. Consider the following example:

Example 3.1.1 Let $X \sim \text{Exp}(1)$ and $Y = 10X$. Then the distribution function of X is of the form

$$F_X(x) = 1 - e^{-x}.$$

We notice that for $u \in (0, \infty)$,

$$\begin{aligned}
\mathbb{P}(X > u \mid Y > u) &= \frac{\mathbb{P}(X > u, Y > u)}{\mathbb{P}(Y > u)} = \frac{\mathbb{P}(X > u, X > \frac{u}{10})}{\mathbb{P}(X > \frac{u}{10})} \\
&= \frac{\mathbb{P}(X > u)}{\mathbb{P}(X > \frac{u}{10})} = \frac{1 - \mathbb{P}(X \leq u)}{1 - \mathbb{P}(X \leq \frac{u}{10})} \\
&= \frac{e^{-u}}{e^{-\frac{u}{10}}} = e^{-\frac{9u}{10}} \quad \underset{u \to \infty}{\longrightarrow} \quad 0.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{P}(Y > u \mid X > u) &= \frac{\mathbb{P}(X > u, Y > u)}{\mathbb{P}(X > u)} = \frac{\mathbb{P}(X > u, X > \frac{u}{10})}{\mathbb{P}(X > u)} \\
&= \frac{\mathbb{P}(X > u)}{\mathbb{P}(X > u)} = 1.
\end{aligned}$$

And so

$$\lim_{u \to \infty} \mathbb{P}(X > u \mid Y > u) \neq \lim_{u \to \infty} \mathbb{P}(Y > u \mid X > u).$$

As Example 3.1.1 shows, the Formula (3.2) is not necessarily symmetric with respect to X and Y. To work around this problem, we apply Lemma 2.1.5 and feed X and Y into their respective distribution functions. Then instead of asking questions about the values of X and Y directly, we ask questions about the probability for the event in which more than an amount u of their probability mass has been covered.

Definition 3.1.2 Let X and Y be random variables and let F_X and F_Y be their respective distribution functions. The *upper tail dependence coefficient* (of X given Y) is defined as the number

$$\lambda_U(X|Y) = \lim_{u \uparrow 1} \mathbb{P}(F_X(X) > u \mid F_Y(Y) > u),$$

when the limit exists. Similarly, the *lower tail dependence coefficient* (of X given Y) is defined as the number

$$\lambda_L(X|Y) = \lim_{u \downarrow 0} \mathbb{P}(F_X(X) \leq u \mid F_Y(Y) \leq u),$$

when the limit exists.

The value $\lambda_U(X|Y)$ is, loosely speaking, the probability for the event that X is "large" when Y is "large". This tells about the dependence of X and Y at the tails of their distributions. When $\lambda_U(X|Y) = 0$ we call X and Y (upper) *tail independent*.

According to Lemma 2.1.5, in the case of continuous distribution functions F_X and F_Y, the random variables $F_X(X)$ and $F_Y(Y)$ are uniformly distributed over $(0, 1)$. Thus in such case

$$\mathbb{P}(F_X(X) > u) = 1 - u = \mathbb{P}(F_Y(Y) > u). \tag{3.3}$$

By utilising the elementary definition for conditional probability the tail dependence coefficient fulfils

$$\begin{aligned}
\lambda_U(X|Y) &= \lim_{u\uparrow 1} \mathbb{P}(F_X(X) > u \mid F_Y(Y) > u) \\
&= \lim_{u\uparrow 1} \frac{\mathbb{P}(F_X(X) > u, F_Y(Y) > u)}{\mathbb{P}(F_Y(Y) > u)} \\
&= \lim_{u\uparrow 1} \frac{\mathbb{P}(F_Y(Y) > u, F_X(X) > u)}{\mathbb{P}(F_X(X) > u)} \\
&= \lambda_U(Y|X).
\end{aligned}$$

Very similarly $\lambda_L(X|Y) = \lambda_L(Y|X)$. For the purposes of this work, we assume the marginal distribution functions to be continuous. This condition could be weakened. We could for example assume the marginals to contain a finite amount of discontinuities, so that after some threshold for u, the Equation (3.3) holds. As far as this work is concerned, the tail dependence coefficients are symmetric with respect to X and Y. For this reason, we may refer to the tail dependence coefficients simply as

$$\lambda_U = \lambda_U(X|Y) = \lambda_U(Y|X) \quad \text{and} \quad \lambda_L = \lambda_L(X|Y) = \lambda_L(Y|X), \tag{3.4}$$

when there is no need to explicitly express which component is being conditioned on.

We investigate how the tail dependence coefficient behaves on academic examples of dependence. We consider pairs of random variables where the marginals are independent and the case where they are almost surely equal.

Example 3.1.3 Let (X, Y) be a random pair with marginal distribution functions F_X and F_Y. Suppose X and Y are independent. Then

$$\lambda_U = \lim_{u\uparrow 1} \mathbb{P}(F_X(X) > u \mid F_Y(Y) > u) = \lim_{u\uparrow 1} \mathbb{P}(F_X(X) > u) = 0$$

and similarly $\lambda_L = 0$.

Suppose then that $X = Y$ almost surely. In this case

$$\begin{aligned}
\lambda_U &= \lim_{u\uparrow 1} \mathbb{P}(F_X(X) > u \mid F_Y(Y) > u) = \lim_{u\uparrow 1} \frac{\mathbb{P}(F_X(X) > u, F_Y(Y) > u)}{\mathbb{P}(F_Y(Y) > u)} \\
&\overset{\text{a.s.}}{=} \lim_{u\uparrow 1} \frac{\mathbb{P}(F_X(X) > u, F_X(X) > u)}{\mathbb{P}(F_X(X) > u)} = 1.
\end{aligned}$$

Again $\lambda_L = 1$ is seen very similarly.

This result is perhaps what should be expected, and it means that at least in terms of these two special cases of dependence structure the tail dependence coefficient works intuitively.

We now compare the tail dependence coefficient and covariance through an example. It is a well-known fact, that uncorrelatedness does not imply independence. In the following, we observe that perhaps unsurprisingly, uncorrelatedness does also not imply tail independence.

Example 3.1.4 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X and Y random variables defined on that probability space, such that, $X \sim \mathcal{N}(0,1)$ and $Y = |X|$. We notice that

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(X|X|) - 0 \cdot \mathbb{E}(|X|) \\
&= \int_\Omega X|X| \mathrm{d}\mathbb{P} \\
&= \int_\Omega X^2 \mathbb{1}\{X \geq 0\} \mathrm{d}\mathbb{P} + \int_\Omega X|X|\mathbb{1}\{X < 0\} \mathrm{d}\mathbb{P} \\
&= \int_\Omega X^2 \mathbb{1}\{X \geq 0\} \mathrm{d}\mathbb{P} + \int_\Omega -X^2 \mathbb{1}\{X \geq 0\} \mathrm{d}\mathbb{P} \\
&= 0,
\end{aligned}$$

since X is symmetrically distributed over 0. Thus we have shown that X and Y are uncorrelated.

Next we show that X and Y are upper tail dependent. We derive a presentation for the distribution function of Y by using the distribution function of X. Note that for any $x \in \mathbb{R}$,

$$\begin{aligned}
F_X(-x) &= \mathbb{P}(X \leq -x) = \mathbb{P}(-X \geq x) = 1 - \mathbb{P}(-X < x) \\
&= 1 - \mathbb{P}(X < x) \qquad (X \text{ is symmetrically distributed over } 0.) \\
&= 1 - \mathbb{P}(X \leq x) \qquad (X \text{ is continuously distributed.}) \\
&= 1 - F_X(x).
\end{aligned} \tag{3.5}$$

Notice also that for any $y \geq 0$,

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(|X| \leq y) = \mathbb{P}(-y \leq X \leq y) \\
&= \mathbb{P}(X \leq y) - \mathbb{P}(X \leq -y) = F_X(y) - F_X(-y) \geq 0.
\end{aligned} \tag{3.6}$$

Fix $\omega \in \Omega$ and let $u \in (0,1)$, such that $F_Y(Y(\omega)) > u$. Now the equation (3.6) results in

$$\begin{aligned}
F_Y(Y(\omega)) &= F_X(Y(\omega)) - F_X(-Y(\omega)) = F_X(|X(\omega)|) - F_X(-|X(\omega)|) \\
&\overset{(3.5)}{=} F_X(|X(\omega)|) - 1 + F_X(|X(\omega)|) = 2F_X(|X(\omega)|) - 1 > u.
\end{aligned} \tag{3.7}$$

By manipulating the last inequality we see that (3.7) implies

$$F_X(|X(\omega)|) > \frac{1+u}{2}. \tag{3.8}$$

Notice also that since X is standard normally distributed, its distribution function fulfills

$$F_X(0) = \frac{1}{2}, \tag{3.9}$$

which means that if F_X maps a number to something greater or equal to $1/2$, that number is positive. By applying (3.9) to (3.8) we obtain

$$F_X(X(\omega)) > \frac{1+u}{2}. \tag{3.10}$$

Now we have shown that (3.7) implies (3.10). This reasoning can be applied in reverse and we notice that

$$\{\omega' \in \Omega \mid F_Y(Y(\omega')) > u\} = \left\{\omega' \in \Omega \mid F_X(X(\omega')) > \frac{1+u}{2}\right\}. \tag{3.11}$$

Now it only remains for us to determine that for $u \in (0,1)$,

$$u < \frac{1+u}{2}. \tag{3.12}$$

Thus,

$$\begin{aligned}
\lambda_U(X|Y) &= \lim_{u\uparrow 1} \mathbb{P}(F_Y(Y) > u \mid F_X(X) > u) \\
&= \lim_{u\uparrow 1} \frac{\mathbb{P}(F_Y(Y) > u \ , \ F_X(X) > u)}{\mathbb{P}(F_X(X) > u)} \\
&\overset{(3.11)}{=} \lim_{u\uparrow 1} \frac{\mathbb{P}(F_X(X) > \frac{1+u}{2} \ , \ F_X(X) > u)}{\mathbb{P}(F_X(X) > u)} \\
&\overset{(3.12)}{=} \lim_{u\uparrow 1} \frac{\mathbb{P}(F_X(X) > \frac{1+u}{2})}{\mathbb{P}(F_X(X) > u)} \\
&\overset{2.1.5}{=} \lim_{u\uparrow 1} \frac{1 - \frac{1+u}{2}}{1-u} = \lim_{u\uparrow 1} \frac{2-(1+u)}{2(1-u)} \\
&= \frac{1}{2}.
\end{aligned}$$

So we have shown that X and Y are tail dependent.

This example illustrates how covariance is not a sufficient measurement of dependence when considering tail events. However, as can be seen from the following example, tail dependence coefficients can also behave against intuition.

Example 3.1.5 Let

$$(X,Y) \sim \mathcal{N}_2(\mathbf{0}, \Sigma), \quad \text{where } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

and

$$\rho = \frac{\mathrm{Cov}\,(X,Y)}{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)} = \mathrm{Cov}\,(X,Y) \in (-1,1),$$

is the correlation. We aim to show that irrespective of the dependence between X and Y they are tail independent. That is,

$$\lambda_U = 0.$$

We remember that the marginals of Gaussian random vectors are normally distributed. Later, Lemma 3.4.1 tells us, that the tail dependence coefficient is not dependent on the marginals. Thus the general case, that is, the case of any bivariate Gaussian distribution corresponds to the one presented in this example.

Proof. We utilise two general facts:

(i) Denote the standard normal tail distribution function by $\overline{\Phi}$ and the standard normal density by φ. We observe, that the ratio

$$\frac{\overline{\Phi}(t)}{\varphi(t)/t} \to 1, \quad \text{as } t \to \infty,$$

or in o-notation,

$$\overline{\Phi}(t) = (1 + o(1))\frac{\varphi(t)}{t}, \quad \text{where } o(1) \to 0, \text{ as } t \to \infty.$$

(ii) The regular conditional distribution of Y given X is known to be

$$Y|X = x \sim \mathcal{N}(x\rho, 1 - \rho^2).$$

Proof of (i): This follows directly from the bounds for the Normal Mill's ratio

$$\frac{t}{t^2 + 1} \leq \frac{\overline{\Phi}(t)}{\varphi(t)} \leq \frac{1}{t}.$$

This can be proven, for example, by differentiating the middle term three times and using various estimates for the density and tail functions. The proof is a bit laborious, but not particularly enlightening, so it is omitted here. It can be found, for example, in [11].

Proof of (ii): Write $Z := Y - \rho X$. Using the bilinearity of covariance

$$\begin{aligned}
\mathrm{Cov}\,(Z, X) &= \mathrm{Cov}\,(Y - \rho X, X) = \mathrm{Cov}\,(Y, X) - \rho\mathrm{Cov}\,(X, X) \\
&= \frac{\mathrm{Cov}\,(Y, X)}{1 \cdot 1} - \rho \cdot 1 = \frac{\mathrm{Cov}\,(Y, X)}{\mathrm{Var}\,(X) \cdot \mathrm{Var}\,(Y)} - \rho \\
&= 0.
\end{aligned}$$

The above says that Z and X are uncorrelated. Since (Z, X) is a linear transformation of (X, Y), it also has a bivariate Gaussian distribution, meaning that the uncorrelatedness of Z and X implies $X \perp\!\!\!\perp Z$. Denote again the standard normal density with φ, denote the joint density of X and Y with $f_{(X,Y)}$, and denote the density of $Z + \rho x$ with $f_{Z+\rho x}$. Then

$$\begin{aligned}
\mathbb{P}(Y \leq y \mid X = x) &= \frac{\int_{-\infty}^{y} f_{(X,Y)}(x, t)dt}{\int_{-\infty}^{\infty} f_{(X,Y)}(x, t)dt} \\
&= \frac{\int_{-\infty}^{y} f_{(X,Z+\rho x)}(x, t)dt}{\varphi(x)} \\
&\overset{\perp\!\!\!\perp}{=} \frac{\int_{-\infty}^{y} \varphi(x) f_{Z+\rho x}(t)dt}{\varphi(x)} \\
&= \int_{-\infty}^{y} f_{Z+\rho x}(t)dt \\
&= \mathbb{P}(Z + \rho x \leq y),
\end{aligned}$$

for any $x, y \in \mathbb{R}$. So we have shown that

$$(Y|X = x) \overset{\mathcal{D}}{=} Z + \rho x.$$

We know that linear combinations of the components of Gaussian vectors are normally distributed. This makes Z and further $(Y|X = x)$ normally distributed. We finish by calculating the expected value and variance:

$$\mathbb{E}(Y|X = x) = \mathbb{E}(Z + \rho x) = \mathbb{E}(Y) - \rho\mathbb{E}(X) + \rho x = \rho x.$$

$$\begin{aligned}\mathrm{Var}\,(Y|X = x) &= \mathrm{Var}\,(Z + \rho x) = \mathrm{Var}\,(Y - \rho X)\\ &= \mathrm{Var}\,(Y) + 2(-\rho)\mathrm{Cov}\,(Y, X) + \rho^2\mathrm{Var}\,(X)\\ &= 1 - 2\rho^2 + \rho^2\\ &= 1 - \rho^2.\end{aligned}$$

This proves claim (ii).

With these two pieces of information, we begin to investigate the upper tail dependence coefficient. Since X and Y are marginals of a Gaussian distribution, they are normally distributed. As such their distribution functions F_Y and F_X are continuous and due to being distribution functions, they are also non-decreasing. This makes them surjective onto $(0, 1)$. Therefore

$$\begin{aligned}\lambda_U(Y|X) &= \lim_{u\uparrow 1} \mathbb{P}(F_Y(Y) > u \mid F_X(X) > u)\\ &\overset{2.1.2\ \text{(i)}}{=} \lim_{u\uparrow 1} \mathbb{P}(F_Y(Y) > F_Y(F_Y^{\leftarrow}(u)) \mid F_X(X) > F_X(F_X^{\leftarrow}(u)))\\ &\overset{2.1.4}{=} \lim_{u\uparrow 1} \mathbb{P}(Y > F_Y^{\leftarrow}(u) \mid X > F_X^{\leftarrow}(u))\\ &= \lim_{t\to\infty} \mathbb{P}(Y > t \mid X > t),\end{aligned}$$

Consider the term inside the limit. Suppose $a > 0$. Then we have

$$\begin{aligned}\mathbb{P}(Y > t \mid X > t) &= \frac{\mathbb{P}(Y > t\ ,\ X > t)}{\mathbb{P}(X > t)}\\ &= \frac{\mathbb{P}(Y > t\ ,\ t + a \geq X > t)}{\mathbb{P}(X > t)} + \frac{\mathbb{P}(Y > t\ ,\ X > t + a)}{\mathbb{P}(X > t)}\end{aligned} \tag{3.13}$$

The first term can be evaluated upward

$$\begin{aligned}\frac{\mathbb{P}(Y > t\ ,\ t + a \geq X > t)}{\mathbb{P}(X > t)} &\leq \frac{\mathbb{P}(Y > t\ ,\ t + a \geq X > t)}{\mathbb{P}(t + a \geq X > t)}\\ &= \mathbb{P}(Y > t \mid t + a \geq X > t).\end{aligned} \tag{3.14}$$

As for the second term, we have

$$\begin{aligned}\frac{\mathbb{P}(Y > t\ ,\ X > t + a)}{\mathbb{P}(X > t)} &\leq \frac{\mathbb{P}(X > t + a)}{\mathbb{P}(X > t)}\\ &\overset{\text{(i)}}{=} (1 + o(1))\frac{t}{t + a} \cdot \frac{\varphi(t + a)}{\varphi(t)}\\ &= (1 + o(1)) \exp\left\{\frac{t^2 - (t + a)^2}{2}\right\}\\ &= (1 + o(1)) \exp\{-at\} \exp\{-a^2/2\}\\ &\underset{t\to\infty}{\longrightarrow} 0.\end{aligned} \tag{3.15}$$

Applying both (3.14) and (3.15) to (3.13), we learn that for any $\varepsilon > 0$, there exists a t so that

$$\mathbb{P}(Y > t \mid X > t) \leq \mathbb{P}(Y > t \mid t + a \geq X > t) + \varepsilon. \tag{3.16}$$

This means that when t grows without limits, the estimate (3.16) applies for arbitrarily small $\varepsilon > 0$. Furthermore, since none of the reasoning used to reach (3.16) concerned the value of a, the estimate holds also for arbitrarily small a, that is, when X is arbitrarily close to t. Thus

$$\begin{aligned}\lambda_U(Y|X) &= \lim_{t\to\infty} \mathbb{P}(Y > t \mid X > t) \\ &\overset{(3.16)}{\leq} \lim_{t\to\infty} \mathbb{P}(Y > t \mid X = t) \\ &\overset{\text{(ii)}}{=} \lim_{t\to\infty} \overline{\Phi}\left(\frac{1}{\sqrt{1-\rho^2}}(1-\rho)t\right) \\ &= 0.\end{aligned}$$

Example 3.1.5 shows that a dependent random pair can still be tail independent. Put succinctly, dependence does not imply tail dependence, however, the reverse is true. As we saw in Example 3.1.3 an independent random pair is always tail independent.

3.2 Copulas

As can be seen from Example 3.1.5, just like covariance, the tail dependence coefficients may also equal 0 when calculated for dependant random variables. One may wonder if it is reasonable to assume that any one-value-descriptor would be sufficient for such a task. Perhaps the boiling down of the dependency structure into one number inevitably leaves out some aspect of the relevant information, so that one can always find an example of a random vector with strong dependency between components that the method does not identify.

This creates a problem, for if one does not get confirmation of dependency using one of such single-value tools, one can not simply rule out the possibility of even a strong dependence relationship existing. One approach to fighting such a problem is to use every possible tool one can think of and hope for the best. Another approach would be to describe the dependency relationship in a more intricate way.

Next, we introduce the concept of copulas, which uses the latter approach. Whereas the tail dependence coefficients simplify the information of dependency into a number, a copula is a function that describes the relationship at any (relevant) point. Thus, the copula leaves out no information.

Definition 3.2.1 Let $d \geq 2$. A d-dimensional *copula* is a distribution function $C : (0,1)^d \to [0,1]$ whose marginal distributions are uniform.

In the case $d = 2$, Definition 3.2.1 can be broken down into a list of three necessary and sufficient conditions. This is very convenient when one has to check and see if a given function $[0,1]^2 \to [0,1]$ is a copula.

Lemma 3.2.2 *A function $C\colon [0,1]^2 \to [0,1]$ is a 2-dimensional copula if and only if it fulfils all the following conditions (i)-(iii).*

(i) $C(0,u) = C(u,0) = 0 \quad \forall u \in [0,1]$.

(ii) $C(1,u) = C(u,1) = u \quad \forall u \in [0,1]$.

(iii) $\forall u_1, v_1, u_2, v_2 \in [0,1],\ u_1 \le u_2,\ v_1 \le v_2,$

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \ge 0.$$

Proof. Assume C to be a copula and therefore, by definition, a multivariate probability distribution function. Let U and V be uniformly distributed on $[0,1]$, with C as their joint distribution. Then based on the distributions of U and V, for $u \in [0,1]$

$$C(0,u) = \mathbb{P}(U \le 0,\ V \le u) = 0 = \mathbb{P}(U \le u,\ V \le 0) = C(u,0).$$

So C fulfils (i). Similarly,

$$\begin{aligned} C(1,u) &= \mathbb{P}(U \le 1,\ V \le u) = \mathbb{P}(U \in [0,1],\ V \le u) = \mathbb{P}(V \le u) = u \\ &= \mathbb{P}(U \le u) = \mathbb{P}(U \le u,\ V \le 1) = C(u,1). \end{aligned}$$

And thus C fulfils (ii). Then let $u_1, v_1, u_2, v_2 \in [0,1]$, such that $u_0 \le u_1 \le u_2$ and $v_0 \le v_1 \le v_2$. Utilising the representation of C as the joint distribution function of random variables U and V, we have

$$\begin{aligned} 0 &\le \mathbb{P}((U,V) \in [u_1,u_2] \times [v_1,v_2]) \\ &= \mathbb{P}((U,V) \in [0,u_2] \times [0,v_2]) - \mathbb{P}((U,V) \in ([0,u_2] \times [0,v_1]) \cup ([0,u_1] \times [0,v_2])) \\ &= C(u_2,v_2) - \Big[\mathbb{P}((U,V) \in [0,u_2] \times [0,v_1]) + \mathbb{P}((U,V) \in [0,u_1] \times [v_1,v_2])\Big] \\ &= C(u_2,v_2) - \Big[C(u_2,v_1) + \big(\mathbb{P}((U,V) \in [0,u_1] \times [0,v_2]) - \mathbb{P}((U,V) \in [0,u_1] \times [0,v_1])\big)\Big] \\ &= C(u_2,v_2) - \Big[C(u_2,v_1) + \big(C(u_1,v_2) - C(u_1,v_1)\big)\Big] \\ &= C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1). \end{aligned}$$

That is C fulfils (iii).

Conversely if C is a function fulfilling conditions (i),(ii) and (iii), it is a function from the unit square to the interval $[0,1]$. By (iii), C is 2-increasing. Based on these two properties, C defines a 2-dimensional distribution function. The boundary conditions (i) and (ii) imply that the marginals are both the distribution function of $\mathcal{U}(0,1)$. □

In order to understand why copulas are a subject of interest for us, we consider the following scenario: Let X_i be random variables with continuous distribution functions $F_{X_i},\ i = 1,\dots,d$. In this context Lemma 2.1.5 tells us that $F_{X_i}(X_i)$ is uniformly distributed over the interval $(0,1)$ for every $i = 1,\dots,d$. Thus the joint distribution of $F_{X_i}(X_i)$ is some d-dimensional copula:

$$C(u_1,\dots,u_d) = \mathbb{P}(F_{X_1}(X_1) \le u_1,\ \dots\ ,F_{X_d}(X_d) \le u_d). \tag{3.17}$$

Let's denote the joint distribution function of $X_i,\ i = 1,\dots,d$, by F. Then by applying the result of Lemma 2.1.4,

$$\begin{aligned} F(x_1,\dots,x_d) &= \mathbb{P}(X_1 \le x_1,\ \dots\ ,X_d \le x_d) \\ &= \mathbb{P}(F_{X_1}(X_1) \le F_{X_1}(x_1),\ \dots\ ,F_{X_d}(X_d) \le F_{X_d}(x_d)) \end{aligned}$$

$$= C(F_{X_1}(x_1),\ \dots\ , F_{X_d}(x_d)). \tag{3.18}$$

So in this case, the copula gives an expression for the joint distribution function using the marginal distributions. While the Equality (3.17) requires continuity from the marginal distribution functions, it turns out that given any random vector, irrespective of the marginal distributions, there exists a copula which fulfils the Equality (3.18). In the case of continuous marginal distributions functions, this copula is also unique. The following theorem is known as Sklar's theorem.

Theorem 3.2.3 *Let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a random vector with marginal distribution functions $F_i,\ i = 1, \dots, d$ and a joint distribution function F. There exists a copula C for which*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \tag{3.19}$$

If functions $F_i,\ i = 1, \dots, d$ are continuous, C is unique.

Proof. The case with continuous marginals we already proved in (3.18) and the unique copula is given by (3.17). In this case, the copula is the joint distribution function of $(F_{X_1}(X_1), \dots, F_{X_d}(X_d))$. There exist multiple different approaches for a proof in the general case. One of them can be found in [3]. □

Given the marginal distributions of a random vector, a copula that fulfils Equation (3.19) gives a way of combining the marginals with each other in a way that produces the joint distribution function. In this way, the copula describes the dependence relationships between the components of the associated random vector.

The applications featured in this work contain distributions with continuous marginals. According to Sklar's theorem then, the copulas which appear in these applications are unique. For this reason we may refer to one of them individually as *the* copula of a random pair.

It is necessary to make note of the fact that copulas do not introduce new information. As opposed to the one-value descriptors of dependence spoken of in Section 3.1, the copula doesn't simplify the dependence structure in any way, rather, the copula contains the full dependence structure itself.

Given the joint distribution function of any random vector, Sklar's theorem promises us that it can be represented using its marginal distribution functions and a copula. This relationship works both ways, in that, given marginal distributions and a copula one can construct the joint distribution of a random vector.

Even so, given a joint distribution function, it can be practically impossible to arrive at an expression for a relevant copula in closed form, even in some very simple cases. Therefore, modelling with copulas often necessarily starts with a decision on the copula and marginals, rather than solving them from a chosen joint distribution.

Since we are interested in the upper tail dependence coefficient λ_U, which particularly concerns survival probabilities, it is of note that the dependence structure of a joint survival distribution function is also a copula. Consider the 2-dimensional case:

Definition 3.2.4 Let (X, Y) be a random pair, with marginal distribution functions F_X and F_Y and a joint distribution function $F_{(X,Y)}$. Denote the copula of (X, Y) by C. Then the *survival copula* of (X, Y) is defined as

$$\widetilde{C}(u, v) := u + v - 1 + C(1 - u, 1 - v),$$

for $u, v \in [0, 1]$.

Notice that for $u \in [0,1]$

$$\widetilde{C}(0,u) = 0 + u - 1 + C(1, 1-u) = u - 1 + 1 - u = 0 = \widetilde{C}(u,0)$$

and

$$\widetilde{C}(1,u) = 1 + u - 1 + C(0, 1-u) = u = \widetilde{C}(u,1).$$

Also for $u_1, v_1, u_2, v_2 \in [0,1]$, so that $u_1 \le u_2$ and $v_1 \le v_2$,

$$\begin{aligned}
\widetilde{C}(u_2,v_2) - \widetilde{C}(u_2,v_1) - \widetilde{C}(u_1,v_2) + \widetilde{C}(u_1,v_1)& \\
&= u_2 + v_2 - 1 + C(u_2,v_2) - u_2 - v_1 + 1 - C(u_2,v_1) \\
&\quad - u_1 - v_2 + 1 - C(u_1,v_2) + u_1 + v_1 - 1 + C(u_1,v_1) \\
&= C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \\
&\ge 0,
\end{aligned}$$

since C is a copula. And so $\widetilde{C}$ fulfils the conditions of Lemma 3.2.2 and thus is also a copula. For the random pair (X,Y) and points $x, y \in R$, we now have the connection

$$\begin{aligned}
\overline{F}_{(X,Y)} &= \mathbb{P}(X > x,\ Y > y) = 1 - \mathbb{P}\left(\{X > x,\ Y > y\}^c\right) \\
&= 1 - \mathbb{P}(X \le x) - \mathbb{P}(Y \le y) + \mathbb{P}(X \le x, Y \le y) \\
&= 1 - \mathbb{P}(X \le x) + 1 - \mathbb{P}(Y \le y) - 1 + C(F_X(x), F_Y(y)) \\
&= \mathbb{P}(X > x) + \mathbb{P}(Y > y) - 1 + C(1 - (1 - F_X(x)), 1 - (1 - F_Y(y))) \\
&= \overline{F}_X(x) + \overline{F}_Y(y) - 1 + C(1 - \overline{F}_X(x), 1 - \overline{F}_Y(y)) \\
&= \widetilde{C}(\overline{F}_X(x), \overline{F}_Y(y)).
\end{aligned} \tag{3.20}$$

In other words, given a random pair and their copula, the survival copula, as defined by 3.2.4 maps the marginal survival distribution functions to the joint survival distribution function at every point $(x,y) \in \mathbb{R}^2$.

The set of all copulas is bounded from above and below by the so-called *Fréchet-Hoeffding bounds* for copulas. We present the 2-dimensional case.

Lemma 3.2.5 *For any copula 2-dimensional copula C and $u, v \in [0,1]$,*

$$\max\{u + v - 1, 0\} \le C(u,v) \le \min\{u, v\}.$$

Proof. Let $u, v \in [0,1]$. Then consider the upper bound first. Condition (iii) of Lemma 3.2.2 implies that C is increasing with respect to both components. With respect to the first component, this can be seen for example by choosing $v_1 = 0$, in which case (iií) gives

$$C(u_2,v_2) - C(u_1,v_2) \ge 0,$$

for any $u_1, u_2, v_2 \in [0,1]$, with $u_1 \le u_2$. The case concerning the second component is similar. Using this, and condition (ii) of Lemma 3.2.2,

$$C(u,v) \le C(1,v) \le v \quad \text{and} \quad C(u,v) \le C(u,1) \le u.$$

Since both upper bounds apply on the whole domain, the copula will always be limited under the smaller component, i.e.

$$C(u,v) \le \min\{u,v\}.$$

For the lower bound, we note that by choosing $u_2 = v_2 = 1$, $u_1 = u$ and $v_1 = v$ the condition (iii) of Lemma 3.2.2 becomes

$$1 - v - u + C(u, v) \geq 0 \quad \Rightarrow \quad C(u, v) \geq u + v - 1.$$

This lower bound goes below zero for some choices of u and v, but we know the does not attain negative values. (By definition the codomain is $[0, 1]$.) Therefore the lower bound is

$$C(u, v) \geq \max\{u + v - 1, 0\}.$$

$\square$

It is worthy of note that the Fréchet-Hoeffding bounds define copulas themselves. We will study the upper bound further in the following section.

3.3 Examples of copulas

A large variety of copulas and families of copulas have become established in use to the extent that they have been given names. We present examples of some commonly encountered copulas which we will utilise later in the simulation study of Section 5. The most simple, and perhaps academic, example is the *Independence copula*:

Example 3.3.1 Given independent real random variables X and Y with marginal distribution functions F_X and F_Y and the joint distribution $F_{(X,Y)}$, we know that independence implies

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y),$$

for every $(x, y) \in \mathbb{R}^2$. Thus a copula fulfilling the equation (3.19), is

$$C(u, v) := uv,$$

since then

$$C(F_X(x), F_Y(y)) = F_X(x)F_Y(y) = F_{(X,Y)}(x, y).$$

Because of this the copula $C(u, v) = uv$ is referred to as the (2-dimensional) Independent copula. The graph and contour plot of the Independent copula are drawn in Figure 3.1.

Intuitively, any random pair with continuous marginals (X, Y) whose copula is the Independent copula, exhibits tail independence, that is, the tail dependence coefficients for the pair are equal to zero. (This is shown later in Section 3.4.) On the other hand, the copula corresponding to "full" dependence is referred to as the *Comonotonic copula*. Perhaps interestingly, the Comonotonic copula corresponds to the upper Fréchet-Hoeffding bound introduced in Lemma 3.2.5.

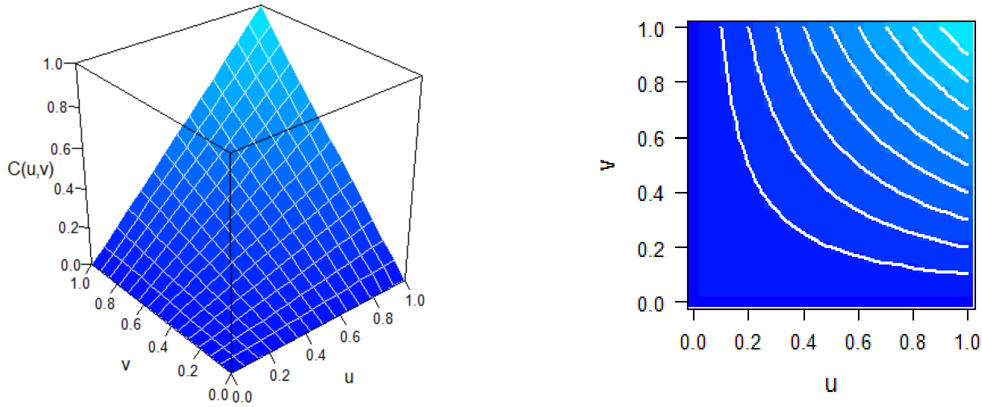
Example 3.3.2 Let (X, Y) be a random pair with joint distribution function $F_{(X,Y)}$ and marginals F_X and F_Y. Suppose $X = Y$ almost surely. As a consequence, they are also equal in distribution, meaning $F_X = F_Y$. Using this we get

$$\begin{aligned} F_{(X,Y)}(x, y) &= \mathbb{P}(X \leq x, Y \leq y) \overset{\text{a.s}}{=} \mathbb{P}(X \leq x, X \leq y) \\ &= \mathbb{P}(X \leq \min\{x, y\}) \overset{F_X \text{ is inc.}}{=} \min\{F_X(x),\ F_X(y)\} \\ &= \min\{F_X(x),\ F_Y(y)\} = C(F_X(x), F_Y(y)), \end{aligned}$$

when we set

$$C(u, v) := \min\{u, v\}.$$

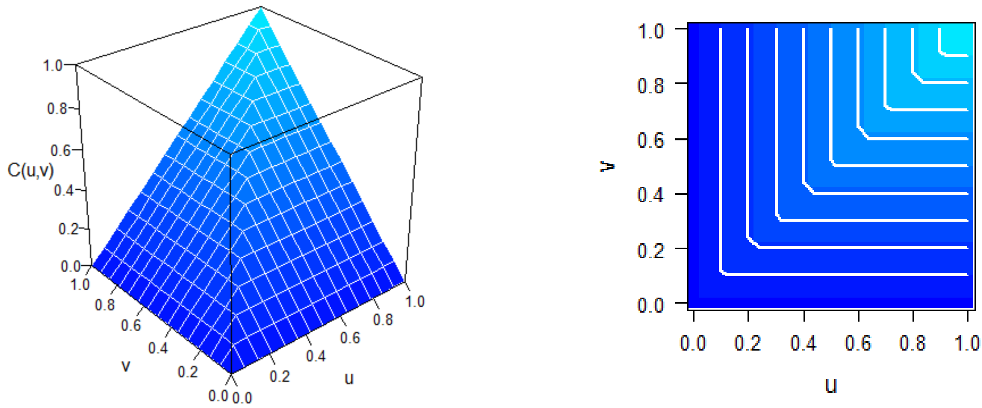
The copula C is referred to as the (2-dimensional) Comonotonic copula and random pairs (X, Y) with the copula C are said to be comonotonic with respect to each other. The graph and contour plot of (an approximation of) the Comonotonic copula are featured in Figure 3.2.



(a) Graph of the Independence copula.

(b) Contour plot of the Independence copula.

Figure 3.1: Plots of the Independence copula.



(a) Graph of an approximation of the Comonotonic copula.

(b) Contour plot of an approximation of the Comonotonic copula.

Figure 3.2: Plots of an approximation of the Comonotonic copula. (The approximation is performed by utilising the Gumbel copula for a large value of the parameter $\theta \geq 1$ (see 3.3.4). For the true Comonotonic copula, the corners of the white contour lines in plot (b) are perfect right angles.)

Next, we consider methods for generating copulas. These methods form classes of copulas, some of which intersect each other. A common class is *Archimedean copulas*. We introduce the 2-dimensional case.

Example 3.3.3 Let $\varphi : [0,1] \to [0,\infty]$ be a strictly decreasing function with $\varphi(1) = 0$. Then the function $(0,1)^2 \to [0,1]$ defined by

$$C(u,v) := \varphi^{\leftarrow}(\varphi(u) + \varphi(v)),$$

satisfies the definition of a copula. (For proof of this, see Theorem 4.1.4 of [15].) A copula that is constructed in this manner is referred to as an *Archimedean copula* and the function φ is called its generator.

We present two examples of parametric families of Archimedean copulas. We are mainly interested in the 2-dimensional cases, but generalisation into higher dimensions is rather intuitive.

Example 3.3.4 The *Frank* family is a parametric family of Archimedean copulas, whose copula and generator are of the form

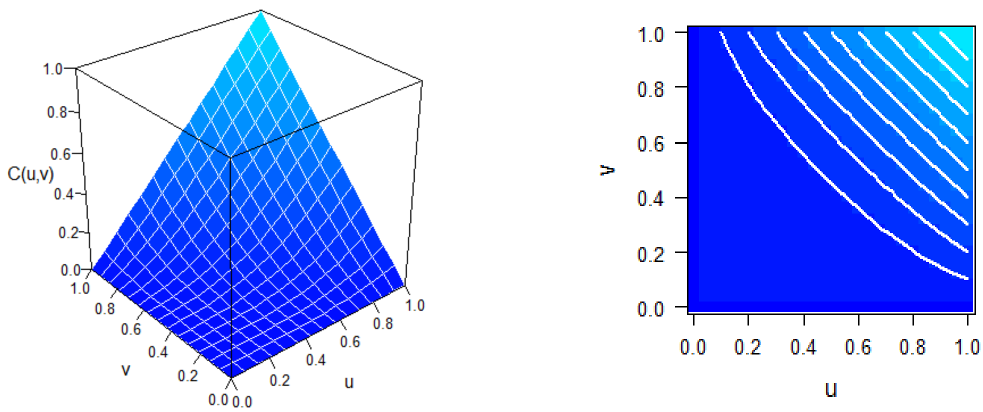
$$C(u,v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right), \quad \varphi(t) = -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right),$$

where the parameter $\theta \in \mathbb{R} \setminus \{0\}$.

Likewise, the *Gumbel* family is another parametric family of Archimedean copulas. The Gumbel copula and generator are of the form

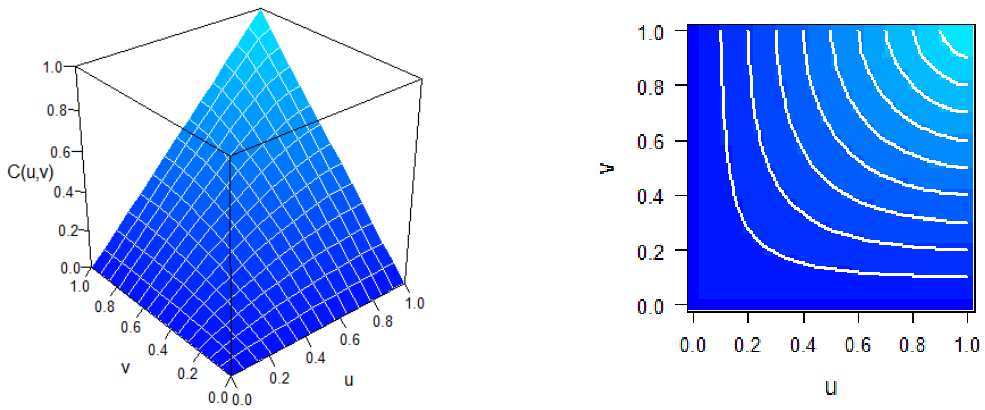
$$C(u,v) = \exp\left\{-\left(\,(-\log(u))^\theta + (-\log(v))^\theta\,\right)^{1/\theta}\right\}, \quad \varphi(t) = (-\log(t))^\theta,$$

with $\theta \in [1,\infty)$.



(a) Graph of a Frank copula. (b) Contour plot of a Frank copula.

Figure 3.3: Plots of the Frank copula with $\theta = -5$. The choice of parametrisation has been made to exaggerate the shape of the graph.



(a) Graph of a Gumbel copula. (b) Contour plot of a Gumbel copula.

Figure 3.4: Plots of the Gumbel copula with $\theta = 1.5$. The choice parametrisation has been made to exaggerate the shape of the graph.

Archimedean copulas are common in application because they describe a relatively large variety of dependence structures.

Given a joint distribution of a random variable, it can be practically impossible to find an expression for the relevant copula in closed form. However, for Gaussian vectors we have the class of *Gaussian copulas*:

Example 3.3.5 Let $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, d$. Then the distribution function of X_i is

$$\Phi_{X_i}(x) = \Phi\left(\frac{x - \mu_i}{\sigma_i}\right), \tag{3.21}$$

where Φ is the standard normal distribution function. Denote

$$\Sigma_{ij} := \mathrm{Cov}\,(X_i, X_j)\,,$$

and

$$M = \begin{bmatrix} 1 & \Sigma_{12} & \Sigma_{13} & \ldots & \Sigma_{1d} \\ \Sigma_{21} & 1 & \Sigma_{23} & \ldots & \Sigma_{2d} \\ \Sigma_{31} & \Sigma_{32} & 1 & \ldots & \Sigma_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{d1} & \Sigma_{d2} & \Sigma_{d3} & \ldots & 1 \end{bmatrix},$$

that is, M is the covariance matrix of $(X_1, \ldots, X_d)$, except for the diagonal of variances, which is replaced by ones. Then consider the multinormally distributed vector

$$\mathbf{Y} := (Y_i)_{i=1,\ldots,d} \sim \mathcal{N}^d(0, M),$$

whose distribution function we denote by $\Phi_M : \mathbb{R}^d \to [0, 1]$. Then

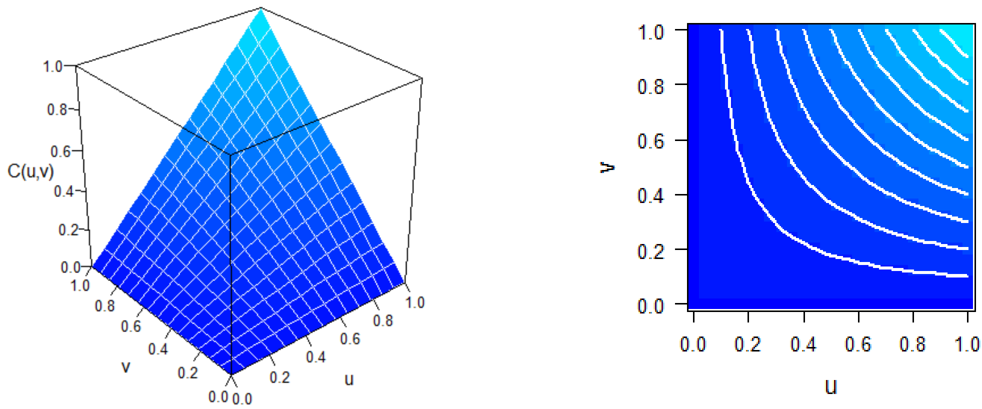
$$(\sigma_1 Y_1 + \mu_1, \ldots, \sigma_d Y_d + \mu_d) \stackrel{\mathcal{D}}{=} (X_1, \ldots, X_d). \tag{3.22}$$

We notice

$$\begin{aligned}
&\Phi_M\left(\Phi^{-1}(\Phi_{X_1}(x_1)),\ \dots\ ,\Phi^{-1}(\Phi_{X_d}(x_d))\right) \\
&\overset{(3.21)}{=} \Phi_M\left(\Phi^{-1}\left(\Phi\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right),\dots,\Phi^{-1}\left(\Phi\left(\frac{x_d-\mu_d}{\sigma_d}\right)\right)\right) \\
&= \Phi_M\left(\frac{x_1-\mu_1}{\sigma_1},\ \dots\ ,\frac{x_d-\mu_d}{\sigma_d}\right) \\
&= \mathbb{P}\left(\sigma_1 Y_1+\mu_1 \le x_1,\ \dots\ ,\sigma_d Y_d+\mu_d \le x_d\right) \\
&\overset{(3.22)}{=} \mathbb{P}(X_1 \le x_1,\dots,X_d \le x_d).
\end{aligned}$$

So the copula, which maps the marginal distribution functions of X_i to the joint Gaussian distribution function of $(X_1,\dots,X_d)$ is

$$C(u_1,\dots,u_d) = \Phi_M(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)).$$



(a) Graph of a Gaussian copula. (b) Contour plot of a Gaussian copula.

Figure 3.5: Plots of a 2-dimensional Gaussian copula where the correlation between components is equal to 0.1. Note the similarity in shape between the contour plot (b) and the contour plot of the Independence copula (b) of Figure 3.1. For larger values of correlation, the plots are more similar to those of the Comonotonic copula in Figure 3.2.

A class of copulas, closely related to tail dependence estimation, is the class of extreme value copulas.

3.3.1 Extreme value copulas

In Section 2.2, we investigated the asymptotic behaviour of maximums of i.i.d. random variables. Now we consider a similar setting, only this time from the viewpoint of copulas. Let

$$\mathbf{X}_j := (X_{j,1},\dots,X_{j,d})\,, \quad j \in \mathbb{N},$$

be a sequence of i.i.d. random vectors. Let F be the distribution function of $\mathbf{X}$ and $F_i,\ i = 1,\ldots,d$ denote the marginal distribution functions. Then denote the component-wise maxima of the first $n \in \mathbb{N}$ elements of the sequence by

$$M_{n,i} := \max\left\{X_{1,i},\ldots,X_{n,i}\right\},$$

, for every $i = 1,\ldots,d$ and

$$\mathbf{M}_n := \left(M_{n,1},\ldots,M_{n,d}\right).$$

Earlier in (2.5), we noticed that the unnormalised margins of $\mathbf{M}_n$ are asymptotically constant and thus we wish to normalise them in a way which is analogical to the normalisation performed in the Theorem 2.2.1. Suppose there exists a sequence $\mathbf{a}_n = (a_{n,1},\ldots,a_{n,d}),\ n \in \mathbb{N}$, where $a_{n,i} > 0$, for $i = 1,\ldots,d$, and a sequence $\mathbf{b}_n = (b_{n,1},\ldots,b_{n,d}),\ n \in \mathbb{N}$, where $b_{n,i} \in \mathbb{R}$, for $i = 1,\ldots,d$ such that the joint distribution function

$$F_{\frac{\mathbf{M}_n-\mathbf{b}_n}{\mathbf{a}_n}}(x_1,\ldots,x_d) := \mathbb{P}\left(\frac{M_{n,1}-b_{n,1}}{a_{n,1}} \le x_1\ ,\ldots,\ \frac{M_{n,d}-b_{n,d}}{a_{n,d}} \le x_d\right) \underset{n\to\infty}{\longrightarrow} \mathbf{G}(x_1,\ldots,x_d),$$

for some non-degenerate distribution function $\mathbf{G} : \mathbb{R}^d \to [0,1]$. We denote the copula associated with the random vector $(\mathbf{M}_n-\mathbf{b}_n)/\mathbf{a}_n$ by $C_{\frac{\mathbf{M}_n-\mathbf{b}_n}{\mathbf{a}_n}}$. Next we investigate the asymptotic properties of this copula.

Perhaps unsurprisingly, we notice that

$$F_{M_{n,i}}(x) = \mathbb{P}(M_{n,i} \le x) = \mathbb{P}\left(\frac{M_{n,i}-b_{n,i}}{a_{n,i}} \le \frac{x-b_{n,i}}{a_{n,i}}\right) = F_{\frac{M_{n,i}-b_{n,i}}{a_{n,i}}}\left(\frac{x-b_{n,i}}{a_{n,i}}\right),$$

for every $i = 1,\ldots,d$ and $x \in \mathbb{R}$. Denote the copula of $\mathbf{M}_n$ by $C_{\mathbf{M}_n}$. Then

$$\begin{aligned} C_{\mathbf{M}_n}(u_1,\ldots,u_d) &= \mathbb{P}\left(F_{M_{n,1}}\left(M_{n,1}\right) \le u_1\ ,\ldots,\ F_{M_{n,d}}\left(M_{n,d}\right) \le u_d\right) \\ &= \mathbb{P}\left(F_{\frac{M_{n,1}-b_{n,1}}{a_{n,1}}}\left(\frac{M_{n,1}-b_{n,1}}{a_{n,1}}\right) \le u_1\ ,\ldots,\ F_{\frac{M_{n,d}-b_{n,d}}{a_{n,d}}}\left(\frac{M_{n,d}-b_{n,d}}{a_{n,d}}\right) \le u_d\right) \\ &= C_{\frac{\mathbf{M}_n-\mathbf{b}_n}{\mathbf{a}_n}}\left(u_1,\ldots,u_d\right). \end{aligned} \tag{3.23}$$

So $\frac{\mathbf{M}_n-\mathbf{b}_n}{\mathbf{a}_n}$ shares a copula with $\mathbf{M}_n$. For this reason, it is sufficient for our purposes to investigate the copula of $\mathbf{M}_n$. Much like in (2.3), we have the identity

$$\begin{aligned} F_{M_{n,i}}(x) &= \mathbb{P}(M_{n,i} \le x) = \mathbb{P}(\max\{X_{1,i}\ ,\ldots,\ X_{n,i}\} \le x) \\ &= \mathbb{P}(X_{1,i} \le x\ ,\ldots,\ X_{n,i} \le x) \overset{\text{i.i.d.}}{=} \prod_{i=1}^{n} \mathbb{P}(X_i \le x) = (F_i(x))^n, \end{aligned} \tag{3.24}$$

for every $i = 1,\ldots,d$ and $x \in \mathbb{R}$. Similarly, for the joint distribution

$$\begin{aligned} F_{\mathbf{M}_n}(x_1,\ldots,x_d) &= \mathbb{P}\left(M_{n,1} \le x_1\ ,\ldots,\ M_{n,d} \le x_d\right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{n}\left\{X_{j,1} \le x_1,\ldots,X_{j,d} \le x_d\right\}\right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{n}\left\{(X_{j,1}\ ,\ldots,\ X_{j,d}) \in (-\infty,x_1] \times \ldots \times (-\infty,x_d]\right\}\right) \end{aligned}$$

$$\overset{\text{i.i.d.}}{=} \prod_{j=1}^{n} \mathbb{P}\big((X_{j,1}\ ,\dots,\ X_{j,d}) \in (-\infty, x_1] \times \dots \times (-\infty, x_d]\big)$$

$$= (F(x_1, \dots, x_d))^n, \tag{3.25}$$

for all $(x_1, \dots, x_d) \in \mathbb{R}^d$. Thus the copula of $\mathbf{M}_n$ satisfies

$$\begin{aligned} C_{\mathbf{M}_n}\left((F_1(x_1))^n, \dots, (F_d(x_d))^n\right) &\overset{(3.24)}{=} C_{\mathbf{M}_n}\left(F_{M_{n,1}}(x_1), \dots, F_{M_{n,d}}(x_d)\right) \\ &= F_{\mathbf{M}_n}(x_1, \dots, x_d) \\ &\overset{(3.25)}{=} (F(x_1, \dots, x_d))^n \\ &= (C_F(F_1(x_1), \dots, F_d(x_d)))^n . \end{aligned} \tag{3.26}$$

In the last step we named the copula corresponding to the random vector $\mathbf{X}$ after its joint distribution function F. Write $u_i = (F_i(x_i))^n$. By substituting this into the above, we obtain

$$C_{\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n}}(u_1, \dots, u_d) \overset{(3.23)}{=} C_{\mathbf{M}_n}(u_1, \dots, u_d) \overset{(3.26)}{=} \left(C_F(u_1^{1/n}, \dots, u_d^{1/n})\right)^n . \tag{3.27}$$

This motivates the following definition.

Definition 3.3.6 We call C an *extreme value copula* if there exists a copula C_F corresponding to a joint distribution function F, such that,

$$\lim_{n\to\infty} \left(C_F(u_1^{1/n}, \dots, u_d^{1/n})\right)^n = C(u_1, \dots, u_d), \tag{3.28}$$

for every $(u_1, \dots, u_d) \in (0,1)^d$.

In the case where such a distribution function F exists, we say that the copula C_F is in the *domain of attraction* of C.

According to equation (3.27), this definition includes the case where C is the point-wise limit of the sequence of copulas $C_{\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n}}$. We also observe the following equivalent property:

Lemma 3.3.7 *Let C be a d-dimensional copula. C is an extreme value copula if and only if it satisfies the equation*

$$C(u_1, \dots, u_d) = \left(C({u_1}^{1/m}, \dots, {u_d}^{1/m})\right)^m, \tag{3.29}$$

for all $m \in \mathbb{N}$ and $(u_1, \dots, u_d) \in (0,1)^d$.

Proof. Suppose C is an extreme value copula. Let $m \in \mathbb{N}$. Then according to the definition, there exists a copula C_F, such that

$$\begin{aligned} C(u_1, \dots, u_d) &\overset{(3.28)}{=} \lim_{n\to\infty} \left(C_F(u_1^{1/n}, \dots, u_d^{1/n})\right)^n \\ &= \lim_{k\to\infty} \left(C_F(u_1^{1/km}, \dots, u_d^{1/km})\right)^{km} \\ &= \lim_{k\to\infty} \left(\left[C_F\big((u_1^{1/m})^{1/k}, \dots, (u_d^{1/m})^{1/k}\big)\right]^k\right)^m \\ &\overset{(3.28)}{=} \left(C(u_1^{1/m}, \dots, u_d^{1/m})\right)^m . \end{aligned}$$

Suppose then that C is a d-dimensional copula such that (3.29) holds. Trivially, one may choose $C_F = C$, in which case

$$\begin{aligned}\lim_{n\to\infty}\Big(C_F(u_1^{1/n},\dots,u_d^{1/n})\Big)^n &= \lim_{n\to\infty}\Big(C(u_1^{1/n},\dots,u_d^{1/n})\Big)^n\\ &= \lim_{n\to\infty} C(u_1,\dots,u_d)\\ &= C(u_1,\dots,u_d),\end{aligned}$$

that is, C is an extreme value copula. □

Thus Lemma 3.3.7 shows, that the class of extreme value copulas are fully characterised by the property of fulfilling Equation (3.29). A copula with this property is also commonly referred to as being *max-stable*. As shown in this following example, the independence, comonotonic and Gumbel copulas are all examples of copulas with this property.

Example 3.3.8 Let $C_{\perp\!\!\!\perp}, C_{\mathrm{Co}}$ and C_{Gum} be the 2-dimensional independence, comonotonic and Gumbel copula with $\theta \in [1,\infty)$, respectively. Fix $(u,v) \in [0,1]^2$ and $r > 0$. Through an elementary calculation, we obtain

$$C_{\perp\!\!\!\perp}(u,v) = uv = \Big(u^{1/r}v^{1/r}\Big)^r = \Big[C_{\perp\!\!\!\perp}(u^{1/r},v^{1/r})\Big]^r.$$

Since r is strictly positive, $t \mapsto t^{1/r}$ is an increasing function. Therefore it preserves order and so

$$C_{\mathrm{Co}}(u,v) = \min\{u,v\} = \Big(\min\Big\{u^{1/r}v^{1/r}\Big\}\Big)^r = \Big[C_{\mathrm{Co}}(u^{1/r},v^{1/r})\Big]^r.$$

Finally, by another elementary calculation,

$$\begin{aligned}C_{\mathrm{Gum}}(u,v) &= \exp\Big\{-\big(\,(-\log(u))^\theta + (-\log(v))^\theta\,\big)^{1/\theta}\Big\}\\ &= \exp\Big\{-\frac{r}{r}\,\big(\,(-\log(u))^\theta + (-\log(v))^\theta\,\big)^{1/\theta}\Big\}\\ &= \left(\exp\left\{-\left(\frac{1}{r^\theta}\left[(-\log(u))^\theta + (-\log(v))^\theta\,\right]\right)^{1/\theta}\right\}\right)^r\\ &= \left(\exp\left\{-\left(\,(-\log(u^{1/r}))^\theta + (-\log(v^{1/r}))^\theta\,\right)^{1/\theta}\right\}\right)^r\\ &= \Big[C_{\mathrm{Gum}}(u^{1/r},v^{1/r})\Big]^r.\end{aligned}$$

So $C_{\perp\!\!\!\perp}, C_{\mathrm{Co}}$ and C_{Gum} are max-stable or, equivalently, extreme value copulas.

3.3.2 Alternative representation of extreme value copulas

The result of Lemma 3.3.7 admits alternative representations for the extreme value copula to be derived. This is because the max-stable property, that is the Equation (3.29), forms a restriction to the type of function C could be, such that it can be presented in a more simple form.

There are different ways this representation could be approached. In this work, we use the method of Pickands and concentrate on the 2-dimensional case.

Lemma 3.3.9 *Let C be a 2-dimensional extreme value copula. Then it has the representation*

$$C(u,v) = \exp\left\{\log(uv) A\left(\frac{\log(v)}{\log(uv)}\right)\right\},$$

where $A : [0,1] \to [1/2,1]$ is a convex function for which

$$A(0) = A(1) = 1$$

holds and

$$\max\{t\ , 1-t\} \leq A(t) \leq 1,$$

for $t \in (0,1)$. The function A is called the Pickands dependence function.

Proof. Let C be a 2-dimensional extreme value copula. Choose (X,Y) to be a random pair whose marginals are exponentially distributed with parameter 1 and whose survival copula is C. Set

$$A(t) := -\log(C(e^{-(1-t)}, e^{-t})) \quad \Rightarrow \quad C(e^{-(1-t)}, e^{-t}) = e^{-A(t)}.$$

Now for any $r,t > 0$, for which $(r(1-t), rt) \in [0,1]^2$, the joint survival function of X and Y becomes

$$\overline{F}_{(X,Y)}(r(1-t), rt) \overset{(3.20)}{=} C\left(\overline{F}_X(r(1-t)), \overline{F}_Y(rt)\right) = C\left(e^{-r(1-t)}, e^{-rt}\right). \tag{3.30}$$

And because Lemma 3.3.7 states that C is max-stable, we have

$$\begin{aligned} C\left(e^{-r(1-t)}, e^{-rt}\right) &= \left[C\left(e^{-r(1-t)/r}, e^{-rt/r}\right)\right]^r = \left[C\left(e^{-(1-t)}, e^{-t}\right)\right]^r \\ &= e^{-rA(t)}. \end{aligned} \tag{3.31}$$

Then we wish to use the change of variables

$$\begin{aligned} &(r(1-t), rt) = (x,y) \quad \Leftrightarrow \quad x = r - rt \text{ and } y = rt \\ \Leftrightarrow \quad & x + y = r \text{ and } \frac{y}{x+y} = t \quad \Leftrightarrow \quad (r,t) = \left(x+y, \frac{y}{x+y}\right). \end{aligned}$$

Together with the Equations(3.30) and (3.31), this gives

$$\overline{F}_{(X,Y)}(x,y) = e^{-(x+y)A\left(\frac{y}{x+y}\right)}. \tag{3.32}$$

Since

$$C(e^{-x}, e^{-y}) = C(F_X(x), F_Y(y)) = \overline{F}_{(X,Y)}(x,y),$$

for any $x,y > 0$, we have

$$C(u,v) = \overline{F}_{(X,Y)}(-\log(u), -\log(v)),$$

for any $u,v \in (0,1)$. This combined with (3.32) becomes

$$\begin{aligned} C(u,v) &= \exp\left\{\big(\log(u) + \log(v)\big) A\left(\frac{\log(v)}{\log(u) + \log(v)}\right)\right\} \\ &= \exp\left\{\log(uv) A\left(\frac{\log(v)}{\log(uv)}\right)\right\}. \end{aligned} \tag{3.33}$$

A is required to be convex for (3.33) to define a copula. The requirement $A(0) = A(1) = 1$ is there to guarantee the border conditions for the copula (condition (ii) of Lemma 3.2.2). Also, combined with convexity, this gives the upper bound $A(t) \leq 1$, for all $t \in [0, 1]$.

The lower bound for A follows from the upper Fréchet-Hoeffding bound for copulas introduced in Lemma 3.2.5. The direction of the inequality is reversed by the minus. Earlier we set

$$\begin{aligned} A(t) &= -\log(C(e^{-(1-t)}, e^{-t})) \overset{3.2.5}{\geq} -\log(\min\{e^{-(1-t)}, e^{-t}\}) \\ &= -\min\{\log(e^{-(1-t)}), \log(e^{-t})\} = -\min\{-(1-t), -t\} \\ &= \max\{1-t, t\}, \end{aligned}$$

for all $t \in (0, 1)$.

□

All possible graphs of the function A lie in the grey triangle in Figure 3.6. The maximal choice $A(t) = 1$ for every $t \in [0, 1]$, yields the Independence copula:

$$C(u, v) = \exp\{\log(uv) \cdot 1\} = uv.$$

Conversely, the minimal choice $A(t) = \max\{t, 1-t\}$, results in the Comonotonic copula:

$$\begin{aligned} C(u, v) &= \exp\left\{\log(uv) \max\left\{\frac{\log(v)}{\log(uv)}, \frac{\log(u)}{\log(uv)}\right\}\right\} \\ &= \exp\left\{\log(uv)\frac{\max\{\log(v), \log(u)\}}{\log(uv)}\right\} \\ &= \exp\{\max\{\log(v), \log(u)\}\} \\ &\overset{0 \leq u,v \leq 1}{=} \exp\{\log(\min\{v, u\})\} \\ &= \min\{u, v\}. \end{aligned}$$

In Example 3.3.8 we showed that the Gumbel copula is an extreme value copula. Thus it follows from Lemma 3.3.9, that it has a Pickands dependence function representation.

Example 3.3.10 Let C be the 2-dimensional Gumbel copula with parameter $\theta \geq 1$. That is,

$$C(u, v) = \exp\left\{-\left(\, (-\log(u))^\theta + (-\log(v))^\theta \,\right)^{1/\theta}\right\},$$

for $(u, v) \in (0, 1)^2$.

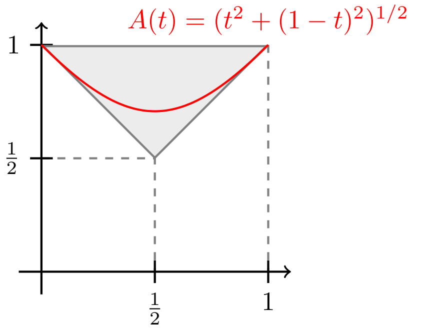
By setting

$$A(t) := \left(t^\theta + (1-t)^\theta\right)^{1/\theta},$$

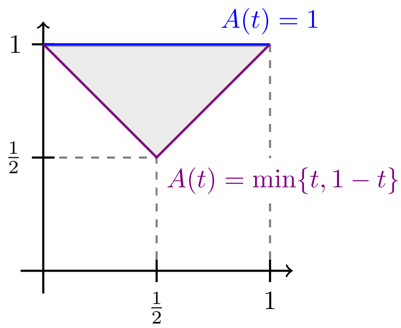
for all $t \in [0, 1]$, we have a function for which

$$\begin{aligned} \max(t, 1-t) &= \left(\max(t^\theta, (1-t)^\theta)\right)^{1/\theta} \leq \left(t^\theta + (1-t)^\theta\right)^{1/\theta} = A(t) \\ &\leq (t + 1 - t)^{1/\theta} = 1, \end{aligned}$$

where the last inequality is true, due to the fact that $0 \leq t \leq 1$ and $\theta \geq 1$. It follows directly from this, that $A(t) \in [1/2, 1]$, for all $t \in [0, 1]$. Clearly, also $A(0) = A(1) = 1$. Thus A is a



(a) The Pickands dependence function of the Gumbel copula, with parameter $\theta = 2$, is drawn in red (see Example 3.3.10).



(b) The Pickands dependence functions of the Independence and the Comonotonic copulas, drawn in blue and purple respectively.

Figure 3.6: All possible graphs of the Pickands dependence function lie inside the grey triangle. It is apparent from (b), that all possible choices of A are bounded by the choices which generate the Independence and Comonotonic copulas.

Pickands dependence function.

We obtain a representation for C with A.

$$\begin{aligned}
A\left(\frac{\log(v)}{\log(uv)}\right) &= \left(\left(\frac{\log(v)}{\log(uv)}\right)^{\theta} + \left(1 - \frac{\log(v)}{\log(uv)}\right)^{\theta}\right)^{1/\theta} \\
&= \left(\frac{\log(v)^{\theta}}{\log(uv)^{\theta}} + \frac{(\log(uv) - \log(v))^{\theta}}{\log(uv)^{\theta}}\right)^{1/\theta} \\
&= \frac{\left(\log(v)^{\theta} + \log(u)^{\theta}\right)^{1/\theta}}{\log(uv)}.
\end{aligned}$$

This implies

$$\begin{aligned}
\exp\left\{(\log(uv))\, A\left(\frac{\log(u)}{\log(uv)}\right)\right\} &= \exp\left\{(\log(uv))\, \frac{\left(\log(u)^{\theta} + \log(v)^{\theta}\right)^{1/\theta}}{\log(uv)}\right\} \\
&= \exp\left\{\left(\log(u)^{\theta} + \log(v)^{\theta}\right)^{1/\theta}\right\} \\
&= \exp\left\{\left((-1)^{\theta}\left((-\log(u))^{\theta} + (-\log(v))^{\theta}\right)\right)^{1/\theta}\right\} \\
&= \exp\left\{-\left((-\log(u))^{\theta} + (-\log(v))^{\theta}\right)^{1/\theta}\right\} \\
&= C(u, v).
\end{aligned}$$

Thus the function A truly does produce the Gumbel copula when it is fed the formula related in Lemma 3.3.9.

3.4 Tail dependence coefficient representation with a limit of the copula

It turns out that the tail dependence coefficients of a random pair can be represented using the pair's copula. Of particular interest to us is the fact that the tail dependence coefficients are not affected by the marginal distributions at all. They are instead only dependent on the diagonal of the relevant copula. This is shown by the following lemma.

Lemma 3.4.1 *Let X and Y be random variables with continuous distribution functions F_X and F_Y respectively. Denote the copula of (X, Y) by C. Then*

$$\lambda_U(Y|X) = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u} \quad \textit{and} \quad \lambda_L(Y|X) = \lim_{u \downarrow 0} \frac{C(u,u)}{u}. \tag{3.34}$$

Proof. Since F_X and F_Y are assumed to be continuous, according to Lemma 2.1.5, we know $F_X(X), F_Y(Y) \sim \mathcal{U}(0,1)$. So for any $u \in (0,1)$, we have

$$\begin{aligned}
\mathbb{P}(F_X(X) > u,\, F_Y(Y) > u) &= 1 - \mathbb{P}(\{F_X(X) \le u\} \cup \{F_Y(Y) \le u\}) \\
&= 1 - \Big[\mathbb{P}(F_X(X) \le u) + \mathbb{P}(F_Y(Y) \le u) - \mathbb{P}(F_X(X) \le u,\ F_Y(Y) \le u)\Big] \\
&= 1 - 2u + C(u,u).
\end{aligned}$$

Consider then the conditional probability featured in the definition of the upper tail dependence coefficient:

$$\begin{aligned}
\mathbb{P}(F_Y(Y) > u \mid F_X(X) > u) &= \frac{\mathbb{P}(F_Y(Y) > u,\ F_X(X) > u)}{\mathbb{P}(F_X(X) > u)} \\
&= \frac{1 - 2u + C(u,u)}{1-u}.
\end{aligned}$$

Similarly for the lower dependence coefficient,

$$\mathbb{P}(F_Y(Y) \le u \mid F_X(X) \le u) = \frac{\mathbb{P}(F_Y(Y) \le u,\ F_X(X) \le u)}{\mathbb{P}(F_X(X) \le u)} = \frac{C(u,u)}{u}.$$

Taking the limits gives us both of the claimed equalities:

$$\lambda_U(Y|X) = \lim_{u \uparrow 1} \mathbb{P}(F_Y(Y) > u \mid F_X(X) > u) = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}$$

and

$$\lambda_L(Y|X) = \lim_{u \downarrow 0} \mathbb{P}(F_Y(Y) \le u \mid F_X(X) \le u) = \lim_{u \downarrow 0} \frac{C(u,u)}{u}.$$

□

As result of Lemma 3.4.1, we have formulas for calculating the tail dependence coefficients based on the copula alone. We can also develop this formula further:

$$2 - \lambda_U \overset{3.4.1}{=} 2 - \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u} = \lim_{u \uparrow 1} \frac{2 - 2u - 1 + 2u - C(u,u)}{1-u}$$

$$= \lim_{u\uparrow 1} \frac{1 - C(u,u)}{1-u} = \lim_{u\uparrow 1} \left[\frac{d}{dt}C(t,t)\right]_{t=u}.$$

For the final equation to hold, the diagonal of the copula is assumed to be differentiable on some interval of the form $(1-\varepsilon, 1)$, where $\varepsilon > 0$. Assuming this is the case, we may simply solve from the above

$$\lambda_U = 2 - \lim_{u\uparrow 1} \left[\frac{d}{dt}C(t,t)\right]_{t=u}. \tag{3.35}$$

This gives a new representation for the tail dependence coefficient. In fact, all the estimators introduced in Section 4 are, in one way or another, based on (3.35). Additionally, it gives a new way of solving the tail dependence coefficient when the copula is sufficiently smooth. We present two examples of this. First, let's solve the upper tail dependence coefficient of the Frank copula.

Example 3.4.2 Let (X,Y) be a random pair whose copula C is the Frank copula with parameter $\theta \in \mathbb{R} \setminus \{0\}$. Then,

$$C(u,v) = -\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right).$$

We wish to calculate the upper tail dependence coefficient using (3.35). Thus we begin by calculating the derivative of the diagonal of this copula. We obtain

$$\begin{aligned}
\frac{d}{dt}C(t,t) &= -\frac{1}{\theta}\frac{d}{dt}\big[\log(t)\big]_{t=1+\frac{(e^{-\theta t}-1)^2}{e^{-\theta}-1}} \frac{d}{dt}\left[1 + \frac{(e^{-\theta t}-1)^2}{e^{-\theta}-1}\right] \\
&= -\frac{1}{\theta}\left(1 + \frac{(e^{-\theta t}-1)^2}{e^{-\theta}-1}\right)^{-1}\left(\frac{1}{e^{-\theta}-1}\right) 2\left(e^{-\theta t}-1\right)(-\theta)e^{-\theta t} \\
&= 2\left(\frac{e^{-\theta}-1}{e^{-\theta}-1+e^{-2\theta t}-2e^{-\theta t}+1}\right)\left(\frac{(e^{-\theta t}-1)e^{-\theta t}}{e^{-\theta}-1}\right) \\
&= 2\left(\frac{e^{-2\theta t}-e^{-\theta t}}{e^{-\theta}+e^{-2\theta t}-2e^{-\theta t}}\right).
\end{aligned}$$

So according to (3.35),

$$\begin{aligned}
\lambda_U &= 2 - \lim_{u\uparrow 1}\left[\frac{d}{dt}C(t,t)\right]_{t=u} \\
&= 2 - \lim_{u\uparrow 1}\left(2\frac{e^{-2\theta u}-e^{-\theta u}}{e^{-\theta}+e^{-2\theta u}-2e^{-\theta u}}\right) \\
&= 2 - 2\left(\frac{e^{-2\theta}-e^{-\theta}}{e^{-2\theta}-e^{-\theta}}\right) \\
&= 0.
\end{aligned}$$

This shows that the Frank copula is upper tail independent for any (permitted) choice of the parameter θ.

In the following (3.35), is used to arrive at an expression for a multivariate t-distributed random pair.

Example 3.4.3 Let $T = (T_1, T_2) \sim \mathbf{t}_\nu^2(\mathbf{0}, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Then

$$\lambda_U = 2 - 2t_{\nu+1}\left(\sqrt{1+\nu}\sqrt{1-\rho}/\sqrt{1+\rho}\right),$$

where $t_{\nu+1}$ is the distribution function of the univariate Student's t distribution with $\nu+1$ degrees of freedom and ρ is the correlation of T_1 and T_2. That is,

$$\rho = \frac{\mathrm{Cov}\,(T_1, T_2)}{\sqrt{\mathrm{Var}\,(T_1)\,\mathrm{Var}\,(T_2)}}.$$

Proof. This proof is divided into two separate claims:

(i) Conditioned on $T_1 = t_1$,

$$\sqrt{\frac{\nu+1}{\nu+t_1^2}}\frac{(T_2 - t_1\rho)}{\sqrt{1-\rho^2}} \sim \mathrm{t}_{\nu+1}.$$

(ii) In this case (and similarly for any radially symmetric copula with a differentiable diagonal), the upper tail dependence coefficient can be calculated as

$$\lambda_U = 2\lim_{x\to\infty}\mathbb{P}(T_2 \geq x \mid T_1 = x),$$

where the equality in the conditional expresses density.

Proof of (i): It follows from the definition of the distribution of T that there exists the representation

$$(T_1, T_2) = T = \frac{X}{Q} = \left(\frac{X_1}{Q}, \frac{X_2}{Q}\right),$$

where $X \sim \mathcal{N}(0, \Sigma)$ and $Q \sim \chi^2_\nu$. It is easy to see from this that the marginals are univariate t-distributed. Particularly

$$T_1 = \frac{X_1}{Q},$$

with $X_1 \sim \mathcal{N}(0,1)$. This follows from the properties of the multivariate normal distribution. This means that

$$T_1 \sim t_\nu.$$

Now consider the conditional probability density function of $T_2 \mid T_1 = t_1$

$$f_{T_2|T_1}(t_2 \mid t_1) = \frac{f_{T_1,T_2}(t_1,t_2)}{f_{T_1}(t_1}.$$

we calculate an expression for the conditional density by substituting in these known density functions:

$$\begin{aligned}
\cdots &= \frac{\frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi\nu\sqrt{\det(\Sigma)}}\left(1+\frac{1}{\nu}(t_1,t_2)^T\Sigma^{-1}(t_1,t_2)\right)^{-\frac{\nu+2}{2}}}{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{t_1^2}{\nu}\right)^{-\frac{\nu+1}{2}}} \\
&= \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi\nu}}\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\nu}{\nu+t_1^2}\right)^{-\frac{\nu+1}{2}}\left(1+\frac{t_1^2-2\rho t_1t_2+t_2^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \\
&= \frac{\Gamma\left(\frac{\nu+2}{2}\right)\sqrt{\nu}}{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi\nu}}\frac{1}{\sqrt{1-\rho^2}\sqrt{\nu+t_1^2}}\left(\frac{\nu}{\nu+t_1^2}\left(1+\frac{t_1^2-2\rho t_1t_2+t_2^2}{\nu(1-\rho^2)}\right)\right)^{-\frac{\nu+2}{2}}
\end{aligned}$$

$$= \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi(\nu+1)}} \frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}\sqrt{\nu+t_1^2}} \left(\frac{\nu}{\nu+t_1^2} + \frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}. \tag{3.36}$$

Denote $\sigma := \sqrt{\frac{\nu+t_1^2}{\nu+1}}\sqrt{1-\rho^2}$ and

$$z = \sigma^{-1}(t_2 - t_1\rho) \Leftrightarrow t_2 = \sigma z + t_1\rho.$$

By substituting this into the right-hand-side term in (3.36), we get

$$\begin{aligned}
&\left(\frac{\nu}{\nu+t_1^2} + \frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \\
&= \left(\frac{\nu}{\nu+t_1^2} + \frac{t_1^2 - 2t_1\rho(\sigma z + t_1\rho) + (\sigma z + t_1\rho)^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \\
&= \left(\frac{\nu}{\nu+t_1^2} + \frac{t_1^2 - 2t_1\rho\sigma z - 2t_1^2\rho^2 + \sigma^2 z^2 + 2t_1\rho\sigma z + t_1^2\rho^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \\
&= \left(\frac{\nu}{\nu+t_1^2} + \frac{(1-\rho^2)t_1^2 + \sigma^2 z^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \\
&= \left(\frac{\nu+t_1^2}{\nu+t_1^2} + \frac{\frac{\nu+t_1^2}{\nu+1}(1-\rho^2)z^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \\
&= \left(1 + \frac{z^2}{\nu+1}\right)^{-\frac{\nu+2}{2}}.
\end{aligned}$$

We also solve the density

$$\frac{dt_2}{dz} = \sigma.$$

Now to prove (i) it suffices to observe that

$$\begin{aligned}
&\mathbb{P}(T_2 \leq y \mid T_1 = t_1) = \mathbb{P}(\sigma^{-1}(T_2 - t_1\rho) \leq \sigma^{-1}(y - t_1\rho) \mid T_1 = t_1) \\
&= \int_{-\infty}^{\sigma^{-1}(y-t_1\rho)} f_{T_2|T_1}(t_2 \mid t_1) dt_2 \\
&= \int_{-\infty}^{\sigma^{-1}(y-t_1\rho)} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi(\nu+1)}} \sigma^{-1} \left(\frac{\nu}{\nu+t_1^2} + \frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{(\nu+t_1^2)(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} dt_2 \\
&\overset{\text{subst.}z}{=} \int_{-\infty}^{\sigma(\sigma^{-1}(y-t_1\rho))+t_1\rho} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi(\nu+1)}} \sigma^{-1}\frac{dt_2}{dz} \left(1 + \frac{z^2}{\nu+1}\right)^{-\frac{\nu+2}{2}} dz \\
&= \int_{-\infty}^{y} \frac{\Gamma\left(\frac{(\nu+1)+1}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\pi(\nu+1)}} \left(1 + \frac{z^2}{\nu+1}\right)^{-\frac{\nu+2}{2}} dz.
\end{aligned}$$

This is the distribution function of a univariate t-distributed random variable with $\nu + 1$ degrees of freedom. Thus claim (i) is proven.

Proof of (ii): Next, we consider the tail dependence coefficient. Represent the copula of the pair (T_1, T_2) by C. We write the diagonal of the copula in a form that makes differentiation more intuitive. Write

$$f : (0,1) \to (0,1)^2, \quad u \mapsto (u,u).$$

Then for all $u \in (0,1)$

$$C(u,u) = (C \circ f)(u).$$

Then we obtain the following using Equality (3.35):

$$\begin{aligned}
\lambda_U &\overset{(3.35)}{=} 2 - \lim_{u\uparrow 1} \left[\frac{d}{ds} C(s,s)\right]_{s=u} = 2 - \lim_{u\uparrow 1} \left[\frac{d}{ds}(C \circ f)(s)\right]_{s=u} \\
&\overset{\text{Chain rule}}{=} 2 - \lim_{u\uparrow 1} \left[\left(\frac{\partial}{\partial x} C(x,y) \ , \ \frac{\partial}{\partial y} C(x,y) \right)^T \Bigg|_{(x,y)=f(s)} \cdot \left(\frac{\partial}{\partial s} s \ , \ \frac{\partial}{\partial s} s \right) \right]_{s=u} \\
&= 2 - \lim_{u\uparrow 1} \left[\frac{\partial}{\partial x} C(x,y) \Bigg|_{(x,y)=f(s)} + \frac{\partial}{\partial y} C(x,y) \Bigg|_{(x,y)=f(s)} \right]_{s=u} \\
&= \lim_{u\uparrow 1} \left[1 - \frac{\partial}{\partial x} C(x,y) \right]_{(x,y)=f(u)} + \lim_{u\uparrow 1} \left[1 - \frac{\partial}{\partial y} C(x,y) \right]_{(x,y)=f(u)} . \qquad (3.37)
\end{aligned}$$

In the following, we will denote the distribution function of a standard Student's t distribution with ν degrees of freedom as t_ν. We continue by using the definition of the copula, the fact that the distribution function t_ν has a true inverse, and the derivative of inverse function formula:

$$\begin{aligned}
1 - \frac{\partial}{\partial x} C(x,y) \Bigg|_{(x,y)=f(u)} &= 1 - \frac{\partial}{\partial x} \mathbb{P}(t_\nu(T_1) \le x \ , \ t_\nu(T_2) \le y) \Bigg|_{(x,y)=f(u)} \\
&= 1 - \frac{\partial}{\partial x} \mathbb{P}(T_1 \le t_\nu^{-1}(x) \ , \ T_2 \le t_\nu^{-1}(y)) \Bigg|_{(x,y)=f(u)} \\
&= 1 - \left[\mathbb{P}(T_1 = t_\nu^{-1}(x) \ , \ T_2 \le t_\nu^{-1}(y)) \frac{1}{t'(t_\nu^{-1}(x))} \right]_{(x,y)=f(u)} \\
&= 1 - \frac{\mathbb{P}(T_1 = t_\nu^{-1}(u) \ , \ T_2 \le t_\nu^{-1}(u))}{P(T_1 = t_\nu^{-1}(u))} \\
&= \mathbb{P}(T_2 \ge t_\nu^{-1}(u) \mid T_1 = t_\nu^{-1}(u))
\end{aligned}$$

where the equal-sign inside of the probability is understood as expressing density. Similarly, the difference inside the right-hand-side limit in (3.37) becomes

$$1 - \frac{\partial}{\partial y} C(x,y) \Bigg|_{(x,y)=f(u)} = \mathbb{P}(T_1 \ge t_\nu^{-1}(u) \mid T_2 = t_\nu^{-1}(u)).$$

The symmetry of the covariance matrix Σ makes the considered distribution radially symmetric. (Showing this is quite straight-forward and to do so one could, for example, explicitly calculate the density function of (T_1, T_2), call it f, and notice that $f(t_1, t_2) = f(t_2, t_1)$.) Because of this radial symmetry and the above calculations, (3.37) takes the form

$$\begin{aligned}
\lambda_U &= \lim_{u\uparrow 1} \mathbb{P}(T_2 \ge t_\nu^{-1}(u) \mid T_1 = t_\nu^{-1}(u)) + \lim_{u\uparrow 1} \mathbb{P}(T_1 \ge t_\nu^{-1}(u) \mid T_2 = t_\nu^{-1}(u)) \\
&= 2 \lim_{u\uparrow 1} \mathbb{P}(T_2 \ge t_\nu^{-1}(u) \mid T_1 = t_\nu^{-1}(u)) = 2 \lim_{x\to\infty} \mathbb{P}(T_2 \ge x \mid T_1 = x).
\end{aligned}$$

Thus (ii) is proven.

Now by combining claims (i) and (ii), it is easy to see that

$$\begin{aligned}
\lambda_U &\overset{\text{(ii)}}{=} 2 \lim_{x\to\infty} \mathbb{P}(T_2 \geq x \mid T_1 = x) \\
&= 2 \lim_{x\to\infty} \mathbb{P}\left(\sqrt{\frac{\nu+1}{\nu+x^2}} \frac{(T_2 - x\rho)}{\sqrt{1-\rho^2}} \geq \sqrt{\frac{\nu+1}{\nu+x^2}} \frac{(x - x\rho)}{\sqrt{1-\rho^2}} \;\middle|\; T_1 = x \right) \\
&\overset{\text{(i)}}{=} 2 - 2 \lim_{x\to\infty} t_{\nu+1} \left(\sqrt{\frac{\nu+1}{\nu+x^2}} \frac{(x - x\rho)}{\sqrt{1-\rho^2}} \right) \\
&= 2 - 2 \lim_{x\to\infty} t_{\nu+1} \left(\sqrt{\frac{\nu+1}{\nu/x^2+1}} \sqrt{\frac{1-\rho}{1+\rho}} \right) \\
&= 2 - 2t_{\nu+1} \left(\sqrt{1+\nu}\sqrt{1-\rho}/\sqrt{1+\rho} \right).
\end{aligned}$$

□

It is possible to develop (3.35) even further in the case that the copula in question is max-stable. In such a case, according to (3.3.9), the copula has a representation using a Pickands dependence function:

$$\begin{aligned}
C(u,v) &= \exp\left\{ \log(uv) A\left(\frac{\log(u)}{\log(uv)} \right) \right\} \\
&= \exp\left\{ \log(u) A\left(\frac{\log(u)}{\log(uv)} \right) + \log(v) A\left(\frac{\log(u)}{\log(uv)} \right) \right\} \\
&= \exp\left\{ \log\left(u^{A\left(\frac{\log(u)}{\log(uv)}\right)} \right) \right\} \exp\left\{ \log\left(v^{A\left(\frac{\log(u)}{\log(uv)}\right)} \right) \right\} \\
&= (uv)^{A\left(\frac{\log(u)}{\log(uv)}\right)}. \qquad (3.38)
\end{aligned}$$

Substituting this into (3.35), we achieve yet another expression for λ_U. The fact that (3.38) is continuously differentiable allows us to get rid of the limit.

$$\begin{aligned}
\lambda_U &= 2 - \lim_{u\uparrow 1} \left[\frac{d}{ds} C(s,s) \right]_{s=u} = 2 - \lim_{u\uparrow 1} \left[\frac{d}{ds} s^{2A\left(\frac{\log(s)}{\log(s^2)}\right)} \right]_{s=u} \\
&= 2 - \frac{d}{du} \left[u^{2A(1/2)} \right]_{u=1} = 2 - 2A(1/2). \qquad (3.39)
\end{aligned}$$

Now we have a simple expression for the (upper) tail dependence coefficient specifically for extreme value copulas. As an example, we calculate the tail dependence coefficient of the Gumbel family, whose Pickands dependence function was introduced in Example 3.3.10.

Example 3.4.4 Let (U,V) be a bivariate random vector with the Gumbel copula C_{Gum} with parameter $\theta \leq 1$ as its distribution function. As shown in Example 3.3.10, the Gumbel copula has a Pickands dependence function representation using the Pickands dependence function

$$A(t) = \left(t^\theta + (1-t)^\theta \right)^{1/\theta}.$$

Now by using the formula (3.39) we get

$$\lambda_U = 2 - 2A(1/2) = 2 - 2\left(\left(\frac{1}{2}\right)^\theta + \left(1 - \frac{1}{2}\right)^\theta \right)^{1/\theta} = 2 - 2^{1/\theta}.$$

4 Estimation

Our goal is to define estimators for the tail dependence coefficient and to investigate their properties. As we have seen in previous sections, the tail dependence coefficient is closely associated with the relevant copula. Two of the estimators presented in Section 4.2 are based solely on the estimation of the underlying copula. For this reason, we introduce an estimator for copulas.

4.1 Empirical copulas

Given a data set, one can obtain information about the hypothetical underlying dependence structure by constructing a pseudo-copula from the data. This object is known as an empirical copula. We present the 2-dimensional case:

Suppose we have $n \in \mathbb{N}$ independent observations of the random vector (X, Y). Denote these observations by $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,n}$. Define the *ranks* of observation $i = 1, \dots, n$ as

$$R_X^{(i)} = \sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq \widetilde{X}_i\} \quad \text{and} \quad R_Y^{(i)} = \sum_{j=1}^{n} \mathbb{1}\{\widetilde{Y}_j \leq \widetilde{Y}_i\}.$$

Then, the *empirical copula* of (X, Y) based on observations $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,n}$ is the same thing as the empirical distribution of (R_X, R_Y) based on observations $(R_X^{(i)}, R_Y^{(i)})_{i=1,\dots,n}$:

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ \frac{R_X^{(i)}}{n} \leq u, \frac{R_Y^{(i)}}{n} \leq v \right\}. \tag{4.1}$$

We then use (4.1) as an estimator for the copula function of the observations. Showing the weak convergence of the empirical copula toward the true copula function, as the sample size is increased, is not a trivial matter. A special case is proven in [20] by utilising the functional delta method. More general cases have subsequently been shown. The proofs, and even the statements of these theorems, are outside the breadth of this thesis.

4.1.1 Observations from an extreme value copula

In order to make use of the properties of extreme value copulas introduced in Section 3.3, one of course needs to be working with an extreme value copula. We need to justify the model assumption, that the data to be analysed is emergent from some extreme value phenomenon. That is, the data points are maximums of a "large" number of observations. Often this is simply not the case. This raises questions about the applicability of the estimators.

Consider random variables X and Y with marginal distribution functions F_X and F_Y and the joint distribution function F. According to Sklar's theorem, there exists a copula C_F, such that,

$$F(x, y) = C_F(F_X(x), F_Y(y)). \tag{4.2}$$

If C_F is in itself an extreme value copula, then it satisfies the max-stability condition (3.29) and thus admits a Pickands dependence function representation. In this case the estimators presented in Section 4.2 can be applied directly to observations from the distribution of (X, Y).

In the case that C_F is not an extreme value copula, we use a method for transforming data into a form approximating extreme value data.

Denote $l \in \mathbb{N}$ observations of (X, Y) with $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,l}$ and consider the component-wise maxima of these observations,

$$\widetilde{X}_l^* = \max\{\widetilde{X}_1, \dots, \widetilde{X}_l\} \quad \text{and} \quad \widetilde{Y}_l^* = \max\{\widetilde{Y}_1, \dots, \widetilde{Y}_l\}.$$

Let C_l^* be the copula fulfilling the equivalent of equation (4.2) for $\widetilde{X}_l^*$ and $\widetilde{Y}_l^*$. Then assuming that the underlying joint distribution function F belongs to the domain of attraction of some extreme value distribution, then according to the reasoning done in Section 3.3, the copula C_l^* belongs to the domain of attraction of an extreme value copula. That is to say, as the amount l of observations $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,l}$ grows, the copula of the component-wise maxima $(\widetilde{X}_l^*, \widetilde{Y}_l^*)$ converges toward an extreme value copula.

Now, for "sufficiently large" l one can treat the observation $(\widetilde{X}_l^*, \widetilde{Y}_l^*)$ as if its distribution is in some way approximate to an extreme value distribution. The *block maxima* and *peak-over-treshold* methods are both ways of transforming data into this maximal form so that extreme-value-theory-specific statistical methods may be applied. We present the method of block maxima.

Suppose we have a sample of $n \in \mathbb{N}$ observations $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,n}$. Seperate the data into m blocks, each with $l = \frac{n}{m}$ observations in them. The number of blocks m is chosen so that the block size l is a whole number. We then replace the observations $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,n}$ with the *block maxima*:

$$\widetilde{X}_j^* = \max\{\widetilde{X}_i \mid (j-1)l < i \leq jl\} \quad \text{and} \quad \widetilde{Y}_j^* = \max\{\widetilde{Y}_i \mid (j-1)l < i \leq jl\}, \tag{4.3}$$

where $j = 1, \dots, m$. The common copula of these m observations is then approximated with the empirical copula

$$\hat{C}_m^*(u, v) = \frac{1}{n} \sum_{j=0}^{n} \mathbb{1}\left\{ \frac{R_{X^*}^{(j)}}{m} \leq u, \frac{R_{Y^*}^{(j)}}{m} \leq v \right\}, \tag{4.4}$$

where $R_{X^*}^{(i)}$ and $R_{Y^*}^{(i)}$ are the ranks of the block maxima $\widetilde{X}_j^*$ and $\widetilde{Y}_j^*$.

The immediate issue applying the method of block maxima creates is the choice of the size of each block (or equivalently the choice of the number of blocks). The block size should be large enough for one to be able to justify applying the reasoning based on asymptotic results related above. On the other hand, to make use of the asymptotic properties of estimators, the number of blocks needs to be large.

Given a sample size n, the larger the block size is, the smaller the number of blocks will be. Even for small block sizes, the sample size will be cut down dramatically. To apply the method of block maxima, one thus needs to be working with considerably large samples to begin with.

The choice of block size also affects the bias and variance of the estimators. Larger block sizes (fewer blocks) lead to smaller estimator biases but larger variances of the estimators. Conversely, a small block size (larger number of blocks) leads to smaller variance of the estimator but a larger estimator bias. The optimal choice of block size for any given situation is a subject of ongoing study. Generally speaking, one wants to choose the block size to be large enough for the resulting estimator to have a small bias and small enough for the estimator to have low variance.

4.2 Estimators for the tail dependence coefficient

We motivate four estimators for the (upper) tail dependence coefficient from different sources. The names of the estimators have been preserved so they are consistent with sources. The estimators are based on equations arrived at in Section 3.4.

4.2.1 Estimators based on the derivative of the copula

In [9], Frahm, Junker and Schmidt introduce two estimators based on the equation (3.35):

$$\lambda_U = 2 - \lim_{u \uparrow 1} \left[\frac{d}{dt} C(t,t) \right]_{t=u}.$$

Consequently, the use of these estimators requires the diagonal of the copula to be continuously differentiable on some open interval whose right-most point is 1.

Suppose we have a sample of size n. An intuitive approach to estimating the derivative of the diagonal of the copula is the difference quotient (or secant) of the empirical copula (4.1) near the point 1. This also motivates the name.

$$\lambda_U^{\text{sec}} := 2 - \frac{1 - \hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right)}{1 - \left(\frac{n-k}{n}\right)}, \tag{4.5}$$

where $0 < k < n$ is the threshold which determines how close to the point 1 the empirical copula is evaluated at.

The second estimator motivated by (3.35), is based on the asymptotic relationship

$$\frac{1 - C(u,u)}{1-u} \approx \frac{\log\left(C(u,u)\right)}{\log(u)},$$

when u is close to 1. Denote

$$\lambda_U^{\log} := 2 - \frac{\log\left(\hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right)\right)}{\log\left(\frac{n-k}{n}\right)}, \tag{4.6}$$

$0 < k < m$. Here, the use of logarithm is motivated by the fact that for completely independent or comonotonic data, $\lambda_U^{\log}$ yields intuitive results independent of the choice of threshold k:

Example 4.2.1 Let $C_{\perp\!\!\!\perp}$ be the Independence copula and C_{Co} the Comonotonic copula. Then for any k and $n \in \mathbb{N}$

$$2 - \frac{\log\left(C_{\perp\!\!\!\perp}\left(\frac{n-k}{n}, \frac{n-k}{n}\right)\right)}{\log\left(\frac{n-k}{n}\right)} = 2 - \frac{\log\left(\left(\frac{n-k}{n}\right)^2\right)}{\log\left(\frac{n-k}{n}\right)} = 2 - \frac{2\log\left(\frac{n-k}{n}\right)}{\log\left(\frac{n-k}{n}\right)} = 0$$

and

$$2 - \frac{\log\left(C_{\text{Co}}\left(\frac{n-k}{n}, \frac{n-k}{n}\right)\right)}{\log\left(\frac{n-k}{n}\right)} = 2 - \frac{\log\left(\frac{n-k}{n}\right)}{\log\left(\frac{n-k}{n}\right)} = 2 - \frac{\log\left(\frac{n-k}{n}\right)}{\log\left(\frac{n-k}{n}\right)} = 1.$$

This shows that the estimator $\lambda_U^{\log}$ should produce intuitive results in the case of independence and comonotony, regardless of the choice of threshold k and sample size n.

Both λ_U^{sec} and $\lambda_U^{\log}$ leave open the choice of threshold k. There is a bias-variance trade-off is associated with k similar to the one associated with the block size of the method of block maxima. The smaller the chosen k is, the lesser the bias $\mathbb{E}(\lambda_U^{\text{sec}}) - \lambda_U$ becomes and the larger the variance will be. Conversely, a larger k will lead to a larger bias, but a smaller variance. The optimal choice of k has not been determined in the general case.

We employ the threshold selection algorithm used by Frahm, Junker and Schmidt in [9]. Their reasoning is the following: The diagonal section of the copula for which the estimator is being calculated is expected to be smooth in some small neighbourhood of the point 1. Additionally, it is expected that the second derivative is small in some such neighbourhood. In such a neighbourhood, the first order derivative must be approximately constant. This means that there must be some interval close to the point 1 on which the map $k \mapsto \hat{\lambda}_U^{\text{sec}}(k)$ is approximately linear.

The goal of this threshold selection algorithm is to observe the behaviour of $\hat{\lambda}_U^{\text{sec}}(k)$ as k grows. This is continued, until the value of $\hat{\lambda}_U^{\text{sec}}(k)$ varies a sufficiently small amount, indicating approximate homogeneity. This is then considered the large enough k to have a small variance for the values of $\hat{\lambda}_U^{\text{sec}}$ and small enough of a k for a small bias.

In practise, we calculate the estimates $\hat{\lambda}_U^{\text{sec}}(k)$ for $k = 1, \dots, n$. We choose a smoothing bandwidth $b \in \mathbb{N}$ and calculate the averages

$$\overline{\lambda_U^{\text{sec}}}(i) = \sum_{k=i}^{i+2b} \hat{\lambda}_U^{\text{sec}}(k),$$

for $i = 1, \dots, n - 2b$. (This is also referred to as employing a box smoothing algorithm on the map $k \mapsto \hat{\lambda}_U^{\text{sec}}(k)$ with a smoothing bandwidth b.) We use the same smoothing bandwidth as is used in [9], where $b = \lfloor 0.005n \rfloor$. This way each average $\overline{\lambda_U^{\text{sec}}}(i)$ consists of approximately 1% of the data. Denote the standard deviation of the averages by

$$\sigma := \sum_{i=1}^{n-2b} \overline{\lambda_U^{\text{sec}}}(i).$$

We then define a plateau length $m = \lfloor \sqrt{n - 2b} \rfloor$ and find the smallest plateau, marked by the index $k = 1, \dots, n - 2n - m + 1$, for which

$$\sum_{i=k+1}^{k+m-1} |\lambda_U^{\bar{\text{sec}}}(i) - \lambda_U^{\bar{\text{sec}}}(k)| \leq 2\sigma. \tag{4.7}$$

The estimator value is then set to the average of the values in the plateau, which is

$$\lambda_U^{\text{sec}} = \frac{1}{m} \sum_{i=1}^{m} \overline{\lambda_U^{\text{sec}}}(k + i - 1).$$

If no plateau fulfils (4.7), the estimator value is set to zero.

This same algorithm is used in threshold selection for the log estimator. The parameters b and m, as well as the plateau condition (4.7), are kept the same to make comparing the performance of the estimators easier.

4.2.2 Estimators based on the Pickands dependence function

Next, we introduce two estimators which utilise the Pickands dependence function representation for an extreme value copula. More particularly they are based on the Equation (3.39):

$$\lambda_U = 2 - 2A(1/2).$$

Genest and Segers derive the following estimator for A in [10]. Suppose we have random variables X and Y with continuous distribution functions F_X and F_Y. Suppose their copula C is max-stable and call its Pickands dependence function A. Denote $U := F_X(X)$ and $V := F_Y(Y)$. Then C is the distribution function of (U, V). We set

$$S := -\log(U) \quad \text{and} \quad T := -\log(V)$$

and investigate the random variable

$$\xi(t) := \min\left\{\frac{S}{1-t}, \frac{T}{t}\right\},$$

when $t \in (0,1)$. Notice that for $x > 0$,

$$\begin{aligned}\mathbb{P}(\xi(t) > x) &= \mathbb{P}(S > (1-t)x \ , \ T > tx) = \mathbb{P}\left(U < e^{-(1-t)x} \ , \ V < e^{-tx}\right) \\ &= \lim_{y\uparrow x} C(e^{-ty}, e^{-(1-t)y}) = \lim_{y\uparrow x}\left(e^{-ty}e^{-(1-t)y}\right)^{A\left(\frac{\log(e^{-ty})}{\log(e^{-ty}e^{-(1-t)y})}\right)} \\ &= \lim_{y\uparrow x} e^{-yA(t)} = e^{-xA(t)}.\end{aligned}$$

That is to say, $\xi(t)$ has an exponential distribution with the rate parameter $A(t)$. Knowing this, we go on to investigate the random variable $-\log(A(t)\xi(t))$. For $x \in \mathbb{R}$, we have

$$\mathbb{P}(-\log(A(t)\xi(t)) \le x) = \mathbb{P}\left(\xi(t) \ge \frac{1}{A(t)}e^{-x}\right) = e^{-A(t)\frac{1}{A(t)}e^{-x}} = e^{-e^{-x}}.$$

Thus, $-\log(A(t)\xi(t))$ has a standard Gumbel distribution. The expected value can be explicitly calculated. First we determine the density function of $-\log(A(t)\xi(t))$:

$$\frac{d}{dx}e^{-e^{-x}} = -e^{-e^{-x}}\frac{d}{dx}(e^{-x}) = -e^{-e^{-x}}(-e^{-x}) = e^{-x-e^{-x}}.$$

Now the mean of the distribution can be calculated as the integral

$$\mathbb{E}(-\log(A(t)\xi(t))) = \int_{-\infty}^{\infty} xe^{-x-e^{-x}}dx. \tag{4.8}$$

Substitute $y = e^{-x}$. Then $x = -\log y$, $\frac{dy}{dx} = -e^{-x}$ and the limits of integration become $\lim_{x\to-\infty} e^{-x} = \infty$ and $\lim_{x\to\infty} e^{-x} = 0$. Thus, (4.8) becomes

$$\int_{-\infty}^{\infty} xe^{-e^{-x}}e^{-x}dx = \int_{\infty}^{0} -\log(y)e^{-y}(-1)dy = \int_{0}^{\infty} \log(y)e^{-y}dy. \tag{4.9}$$

Then, using the fact that

$$\frac{\partial}{\partial t}y^t e^{-y} = y^t \log(y)e^{-y},$$

the last integral in (4.9) can be further simplified into

$$\int_0^{\infty}\left[\frac{\partial}{\partial t}y^t e^{-y}\right]_{t=0} dy = \frac{\partial}{\partial t}\left[\int_0^{\infty} y^t e^{-y}dy\right]_{t=0} = \frac{\partial}{\partial t}\left[\Gamma(t+1)\right]_{t=0} = \Gamma'(1),$$

where Γ is the gamma function. This number is known as the Euler-Mascheroni constant

$$\gamma = \Gamma'(1) \approx 0.5772.$$

Now we have shown

$$\mathbb{E}(-\log(A(t)\xi(t))) = \gamma.$$

And so we attain

$$\log(A(t)) = -\gamma - \mathbb{E}(\log(\xi(t))). \tag{4.10}$$

We now need a way to estimate the expected values involving the random variable $\xi(t)$ from given observations $(\widetilde{X}_i, \widetilde{Y}_i)_{i\in\{1,\dots,n\}}$. Set

$$\hat{U}_i := \frac{1}{1+n}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq \widetilde{X}_i\} \text{ and } \hat{V}_i := \frac{1}{1+n}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{Y}_j \leq \widetilde{Y}_i\},$$

The normalisation is done using $n+1$ instead of n to avoid reaching 1. This, and the observation that $\hat{U}_i, \hat{V}_i > 0$, for every $i = 1, \dots, n$, allows us to not worry about division by zero and attempting to evaluate logarithms at zero in what follows. We also write

$$\hat{S}_i = -\log \hat{U}_i \text{ and } \hat{T}_i = -\log \hat{V}_i.$$

Now define

$$\hat{\xi}_i(t) := \begin{cases} \hat{S}_i, & t = 0 \\ \min\left\{\frac{\hat{S}_i}{1-t}, \frac{\hat{T}_i}{t}\right\}, & t \in (0,1) \\ \hat{T}_i, & t = 1. \end{cases}$$

Using this to estimate $\xi(t)$, Equality (4.10) leads us to the estimator

$$\log \hat{A}_u^{\mathrm{CFG}}(t) = -\gamma - \frac{1}{n}\sum_{i=1}^{n} \log(\hat{\xi}_i(t)).$$

The name of this estimator is due to [2], in which Capéraà, Fougères and Genest introduce an estimator for A, which Segers simplifies further into the above form in [10].

The subindex u in the above estimator denotes the fact that it doesn't define a Pickands dependence function since it doesn't fulfil the property $A(0) = A(1) = 1$. Thus it is the *u*ncorrected version of estimator studied by Genest and Segers in [10]. They performed the correction by setting

$$\log \hat{A}_{ab}^{\mathrm{CFG}}(t) := \log \hat{A}_u^{\mathrm{CFG}}(t) - a(t)\left(\log \hat{A}_u^{\mathrm{CFG}}(0)\right) - b(t)\left(\log \hat{A}_u^{\mathrm{CFG}}(1)\right),$$

where $a, b : [0,1] \to \mathbb{R}$ are continuous mappings chosen so that the boundary conditions for A are met and the estimator is afforded optimal behaviour. Finding optimal a and b is discussed by Segers in [18]. For our purposes we use $a(t) = 1 - t$ and $b(t) = t$, since these seem to perform well according to Segers (see [18]).

Finally, we have settled on the estimator

$$\begin{aligned} \hat{A}^{\mathrm{CFG}}(t) :=& \exp\left\{ \log \hat{A}_u^{\mathrm{CFG}}(t) - (1-t)\left(\log \hat{A}_u^{\mathrm{CFG}}(0)\right) - t\left(\log \hat{A}_u^{\mathrm{CFG}}(1)\right) \right\} \\ =& \exp\left\{ -\gamma - \frac{1}{n}\sum_{i=1}^{n}\log(\hat{\xi}_i(t)) \right. \\ & \left. - (1-t)\left(-\gamma - \frac{1}{n}\sum_{i=1}^{n}\log(\hat{S}_i)\right) - t\left(-\gamma - \frac{1}{n}\sum_{i=1}^{n}\log(\hat{T}_i)\right) \right\} \\ =& \exp\left\{ \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\hat{S}_i^{(1-t)}\hat{T}_i^{t}}{\hat{\xi}_i(t)}\right) \right\}. \end{aligned} \tag{4.11}$$

And so, we have a new estimator for the (upper) tail dependence coefficient as well. Based on the above and (3.39):

$$\begin{aligned}\lambda_U^{\text{CFG}} :=& 2-2\hat{A}^{\text{CFG}}(1/2)\\ =&2-2\exp\left\{\frac{1}{n}\sum_{i=1}^{n}\log\left(\sqrt{\hat{S}_i\hat{T}_i}\Big/ 2\min\left\{\hat{S}_i,\hat{T}_i\right\}\right)\right\}\\ =&2-2\exp\left\{\frac{1}{n}\sum_{i=1}^{n}\log\left(\sqrt{\log\frac{1}{\hat{U}_i}\log\frac{1}{\hat{V}_i}}\Big/ 2\min\left\{\log\frac{1}{\hat{U}_i},\log\frac{1}{\hat{V}_i}\right\}\right)\right\}\\ =&2-2\exp\left\{\frac{1}{n}\sum_{i=1}^{n}\log\left(\sqrt{\log\frac{1}{\hat{U}_i}\log\frac{1}{\hat{V}_i}}\Big/ \log\frac{1}{\max\left\{\hat{U}_i,\hat{V}_i\right\}^2}\right)\right\}. \end{aligned}\tag{4.12}$$

Frahm, Junker and Schmidt also considered the λ_U^{CFG} estimator (4.12) in [9], finding it to generally perform better than estimators λ_U^{sec} (4.5) and $\lambda_U^{\log}$ (4.6). The asymptotic properties of $\hat{A}^{\text{CFG}}(t)$ are investigated in detail by Genest and Segers in [10].

Another estimator for λ_U based on the Pickands dependence function is introduced by Ferreira in [8]. We once more use the notation of two random variables X and Y with continuous distribution functions F_X and F_Y. We denote their (max-stable) copula by C and its Pickands dependence function by A. Write $U := F_X(X)$ and $V := F_Y(Y)$. Then C is the distribution function of (U, V). Ferreiras' estimator is based on the observation that, for $t \in (0,1)$,

$$\begin{aligned}\mathbb{P}(\max\{U^{1/x},V^{1/y}\}\le t) &= \mathbb{P}(U\le t^x, V\le t^y) = C(t^x,t^y)\\ &= (t^x t^y)^{A\left(\log(t^x)\big/[\log(t^x)+\log(t^y)]\right)}\\ &= t^{(x+y)A\left(\frac{x}{x+y}\right)}.\end{aligned}\tag{4.13}$$

We obtain the density by differentiating

$$\frac{d}{dt}\mathbb{P}(\max\{U^{1/x},V^{1/y}\}\le t) = (x+y)A\left(\frac{x}{x+y}\right)t^{(x+y)A\left(\frac{x}{x+y}\right)-1}.$$

The density allows us to calculate the expected value

$$\begin{aligned}\mathbb{E}\left(\max\{U^{1/x},V^{1/y}\}\right) &= \int_0^1 \frac{d}{dt}\mathbb{P}(\max\{U^{1/x},V^{1/y}\}\le t)t\, dt\\ &= \int_0^1 (x+y)A\left(\frac{x}{x+y}\right)t^{(x+y)A\left(\frac{x}{x+y}\right)-1}t\, dt\\ &= (x+y)A\left(\frac{x}{x+y}\right)\int_0^1 t^{(x+y)A\left(\frac{x}{x+y}\right)}dt\\ &= (x+y)A\left(\frac{x}{x+y}\right)\left[\frac{1}{(x+y)A\left(\frac{x}{x+y}\right)+1}t^{(x+y)A\left(\frac{x}{x+y}\right)+1}\right]_0^1 dt\\ &= \frac{(x+y)A\left(\frac{x}{x+y}\right)}{(x+y)A\left(\frac{x}{x+y}\right)+1}.\end{aligned}$$

Now A can be solved from the above:

$$\mathbb{E}\left(\max\{U^{1/x}, V^{1/y}\}\right) = \frac{(x+y)A\left(\frac{x}{x+y}\right)}{(x+y)A\left(\frac{x}{x+y}\right)+1}$$
$$\Leftrightarrow$$
$$\mathbb{E}\left(\max\{U^{1/x}, V^{1/y}\}\right)\left((x+y)A\left(\frac{x}{x+y}\right)+1\right) = (x+y)A\left(\frac{x}{x+y}\right)$$
$$\Leftrightarrow$$
$$(x+y)A\left(\frac{x}{x+y}\right)\left(1-\mathbb{E}\left(\max\{U^{1/x}, V^{1/y}\}\right)\right) = \mathbb{E}\left(\max\{U^{1/x}, V^{1/y}\}\right)$$
$$\Leftrightarrow$$
$$A\left(\frac{x}{x+y}\right) = \frac{\mathbb{E}\left(\max\{U^{1/x}, V^{1/y}\}\right)}{(x+y)\left(1-\mathbb{E}\left(\max\{U^{1/x}, V^{1/y}\}\right)\right)}$$

Since the connection of A to the tail dependence estimator concerns only the value of A at $\frac{1}{2}$, we choose $x = y = 1$ and obtain

$$A\left(\frac{1}{2}\right) = \frac{\mathbb{E}\left(\max\{U, V\}\right)}{2\left(1-\mathbb{E}\left(\max\{U, V\}\right)\right)}. \tag{4.14}$$

The expected values in the above are then estimated with the arithmetic means of maximums of observation vectors. By making this replacement we arrive at the estimator introduced by Ferreira and Ferreira (hence the name)

$$\begin{aligned}
\lambda_U^{\mathrm{FF}} :=& 2 - \frac{2\,\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}}{2\left(1-\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}\right)} \\
=& \frac{2\,\left(1-\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}\right)-\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}}{1-\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}} \\
=& \frac{3-3\,\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}-1}{1-\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}} \\
=& 3-\left(1-\frac{1}{n}\sum_{i=1}^{n}\max\left\{\hat{U}_i, \hat{V}_i\right\}\right)^{-1}.
\end{aligned} \tag{4.15}$$

Both the CFG and FF estimators are based on reasoning, which assumes the distribution underlying the observations to have a max-stable copula. As such, applying them to unaltered non-extreme value data is questionable. In Section 5, we deal with such scenarios by applying the method of block maxima and thus approximating extreme value data with the given non-extreme value data. This is discussed more in Section 4.1, where we introduced the method of block maxima, and Section 5.1, where our choice of the number of blocks is discussed.

The definitions of the estimators contained in this section are compiled into Definition 4.2.2 below.

Definition 4.2.2 Let $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\ldots,n}$ be a sequence of $n \in \mathbb{N}$ i.i.d. random pairs. Denote the empirical copula of these pairs, defined in (4.1), by $\hat{C}_n$. If the common distribution of the observations $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\ldots,n}$ is an extreme value distribution, we define

$$\hat{U}_i := \frac{1}{n+1} \sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq \widetilde{X}_i\} \quad \text{and} \quad \hat{V}_i := \frac{1}{n+1} \sum_{j=1}^{n} \mathbb{1}\{\widetilde{Y}_j \leq \widetilde{Y}_i\}.$$

Otherwise we decide on a block size l, such that the number of blocks $m := n/l \in \mathbb{N}$ and set

$$\hat{U}_i := \frac{1}{m+1} \sum_{j=1}^{m} \mathbb{1}\{\widetilde{X}_j^* \leq \widetilde{X}_i^*\} \quad \text{and} \quad \hat{V}_i := \frac{1}{m+1} \sum_{j=1}^{m} \mathbb{1}\{\widetilde{Y}_j^* \leq \widetilde{Y}_i^*\},$$

where the block maxima $\widetilde{X}_j^*$ and $\widetilde{Y}_j^*$, for $j = 1, \ldots, m$, are defined in (4.3).

Then we define

$$\begin{aligned}
\lambda_U^{\text{sec}} &:= 2 - \frac{1 - \hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right)}{1 - \left(\frac{n-k}{n}\right)} \\
\lambda_U^{\text{log}} &:= 2 - \frac{\log\left(\hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right)\right)}{\log\left(\frac{n-k}{n}\right)} \\
\lambda_U^{\text{CFG}} &:= 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n} \log\left(\sqrt{\log\frac{1}{\hat{U}_i}\log\frac{1}{\hat{V}_i}} \Big/ \log\frac{1}{\max\left\{\hat{U}_i, \hat{V}_i\right\}^2}\right)\right\} \\
\lambda_U^{\text{FF}} &:= 3 - \left(1 - \frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\}\right)^{-1}.
\end{aligned}$$

4.3 Properties of tail dependence estimators

In this section, we continue to use the notation where (X, Y) is a generic representative variable from the sampling distribution. The continuous marginal distribution functions of the sampling distribution are denoted by F_X and F_Y. Continuity is required for the reasoning in Section 4.2 (for example, Equation (4.13)) to still hold. Observations are denoted by $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\ldots,n}$ and the empirical marginal distribution functions by

$$\hat{F}_X(x) := \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq x\} \quad \text{and} \quad \hat{F}_Y(y) := \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{Y}_j \leq y\}.$$

We write $U := F_X(X)$ and $V := F_Y(Y)$ and similarly $\hat{U}_i := \hat{F}_X(\widetilde{X}_i)$ and $\hat{U}_i := \hat{F}_Y(\widetilde{Y}_i)$.

The asymptotic properties of the sec estimator λ_U^{sec} are considered by Schmidt and Stadtmüller in [17]. They prove strong consistency (See Theorem 6 in [17]) and asymptotic normality (See Theorem 5 in [17]) of λ_U^{sec} by applying weak convergence results to the empirical copula process.

Genest and Segers consider the asymptotic properties of the CFG estimator for the Pickands dependence function defined in (4.11). Among other things, they show $\hat{A}^{\text{CFG}}$ to be consistent and asymptotically unbiased. These results are easily extendable to λ_U^{CFG}, as it is an affine transformation of $\hat{A}^{\text{CFG}}$.

In [8], Ferreira shows that the FF estimator λ_U^{FF} is strongly consistent. Also, under some qualifying assumptions, Ferreira and Ferreira show asymptotic normality for their estimator in [7]. We present these results with elaborations on Ferreiras' proofs.

Theorem 4.3.1 *Let $(\widetilde{X}_i, \widetilde{Y}_i)_{i\in\{1,\dots,n\}}$ be a sample of i.i.d. observations from a bivariate distribution with a max-stable copula and continuous marginals.*

When calculated for such data, λ_U^{FF} is strongly consistent. In other words,

$$\lim_{n\to\infty} \lambda_U^{FF} = \lambda_U$$

almost surely.

Proof. Let's investigate the difference

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\} - \mathbb{E}\big(\max\left\{U, V\right\}\big)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{F}_X(\widetilde{X}_i), \hat{F}_Y(\widetilde{Y}_i)\right\} - \mathbb{E}\big(\max\left\{F_X(X), F_Y(Y)\right\}\big)\right| \\
&\leq \left|\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{F}_X(\widetilde{X}_i), \hat{F}_Y(\widetilde{Y}_i)\right\} - \frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\} - \mathbb{E}\big(\max\left\{F_X(X), F_Y(Y)\right\}\big)\right|. \qquad (4.16)
\end{aligned}$$

Consider the latter term in (4.16). We wish to apply the strong law of large numbers, so we have to verify that the random variables involved are i.i.d. and that the expected value of their common distribution indeed exists.

The sequence $\{(\widetilde{X}_i, \widetilde{Y}_i)\}_{i=1,\dots,n}$ is i.i.d. by assumption, and therefore the sequence $\Big\{\max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}\Big\}_{i=1,\dots,n}$ is also i.i.d. The expected value we calculate explicitly. By (4.13), we have, for every $t \in (0,1)$ and $i \in \{1,\dots,n\}$,

$$\begin{aligned}
\mathbb{P}\left(\max\left\{F_X(X_i), F_Y(Y_i)\right\} \leq t\right) &= \mathbb{P}\left(\max\left\{F_X(X), F_Y(Y)\right\} \leq t)\right) \\
&= \mathbb{P}\left(\max\left\{U, V\right\} \leq t\right) \\
&= t^{2A(1/2)}.
\end{aligned}$$

We remember that since $\max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}$ is a non-negative random variable for all $i \in \{1,\dots,n\}$, the expected value for any $i \in \{1,\dots,n\}$ can be determined by integrating the tail probability

$$\begin{aligned}
\mathbb{E}\left(\max\left\{F_X(X_i), F_Y(Y_i)\right\}\right) &= \int_0^1 \mathbb{P}\left(\max\left\{F_X(X_i), F_Y(Y_i)\right\} > t\right) dt \\
&= \int_0^1 1 - t^{2A(1/2)} dt \\
&= 1 - \frac{1}{1 + 2A(1/2)}
\end{aligned}$$

$$= \frac{2A(1/2)}{1+2A(1/2)} < \infty. \tag{4.17}$$

So the expected value is finite. Thus, by the strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\} \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}\big(\max\left\{F_X(X), F_Y(Y)\right\}\big).$$

Moreover,

$$\left|\frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\} - \mathbb{E}\big(\max\left\{F_X(X), F_Y(Y)\right\}\big)\right| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Consider then the former term in (4.16):

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{F}_X(\widetilde{X}_i), \hat{F}_Y(\widetilde{Y}_i)\right\} - \frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}\right| \\
&= \frac{1}{n}\sum_{i=1}^{n} \left|\max\left\{\hat{F}_X(\widetilde{X}_i), \hat{F}_Y(\widetilde{Y}_i)\right\} - \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \left|\hat{F}_X(\widetilde{X}_i) + \hat{F}_Y(\widetilde{Y}_i) - F_X(\widetilde{X}_i) - F_Y(\widetilde{Y}_i)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \left|\hat{F}_X(\widetilde{X}_i) - F_X(\widetilde{X}_i)\right| + \frac{1}{n}\sum_{i=1}^{n} \left|\hat{F}_Y(\widetilde{Y}_i) - F_Y(\widetilde{Y}_i)\right| \\
&\leq \sup_{x\in\mathbb{R}} \left|\hat{F}_X(x) - F_X(x)\right| + \sup_{y\in\mathbb{R}} \left|\hat{F}_Y(y) - F_Y(y)\right|
\end{aligned}$$

By the Glivenko-Cantelli Theorem (Theorem 2.3.2), both

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_X(x) - F_X(x)\right| \xrightarrow[n\to\infty]{\text{a.s.}} 0 \quad \text{and} \quad \sup_{y\in\mathbb{R}} \left|\hat{F}_Y(y) - F_Y(y)\right| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

A combination of our observations shows us that (4.16) converges to zero almost surely. In other words,

$$\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\} \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}\big(\max\left\{U, V\right\}\big).$$

Using this and the equality (4.14), we get

$$\frac{\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\}}{2\left(1 - \frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\}\right)} \xrightarrow[n\to\infty]{\text{a.s.}} \frac{\mathbb{E}\big(\max\left\{U, V\right\}\big)}{2\left(1 - \mathbb{E}\big(\max\left\{U, V\right\}\big)\right)} \overset{(4.14)}{=} A\left(\frac{1}{2}\right).$$

Which means that

$$\lambda_U^{\text{FF}} = 2 - 2\frac{\frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\}}{2\left(1 - \frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\}\right)} \xrightarrow[n\to\infty]{\text{a.s.}} 2 - 2A(1/2) = \lambda_U.$$

□

In [7], Ferreira and Ferreira prove the asymptotic normality of the FF estimator when it is calculated using known marginal distribution functions. That is, instead of estimating the marginals with the empirical distribution functions like in the definition of λ_U^{FF} (4.15), the observations are plugged directly into the known marginals. Define

$$\lambda_U^{\mathrm{FF}^*} := 3 - \left(1 - \frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}\right)^{-1}.$$

Asymptotic normality can now be shown for this version of the FF estimator.

Theorem 4.3.2 *Let $(\widetilde{X}_i, \widetilde{Y}_i)_{i\in\{1,\ldots,n\}}$ be a sample of i.i.d. observations from a bivariate distribution with a max-stable copula. Suppose that the marginal distribution functions F_X and F_Y are continuous and known.*

When calculated for such data, $\lambda_U^{FF^*}$ is asymptotically normal, in the sense that

$$\sqrt{n}(\lambda_U^{FF^*} - \lambda_U) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \frac{A(1/2)(1 + 2A(1/2))^2}{1 + A(1/2)}.$$

Proof. We show convergence first for $\frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}$ and then apply the delta method with a function which transforms this term into the estimator $\lambda_U^{\mathrm{FF}^*}$. We begin with the claim, that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\} - \mathbb{E}\big(\max\left\{U, V\right\}\big)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{\max}), \tag{4.18}$$

for some $\sigma^2_{\max} \geq 0$. To prove this, we apply the central limit theorem. Since by assumption we have an i.i.d. sequence of random variables $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\ldots,n}$, the sequence of maxima $\left\{\max\left\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\right\}\right\}_{i=1,\ldots,n}$ is also i.i.d. Therefore we only need to check that the variance is finite. For this purpose, we calculate the second moment

$$\begin{aligned}\mathbb{E}\big(\max\left\{U, V\right\}^2\big) &= \int_0^1 1 - \mathbb{P}(\max\left\{U, V\right\} \leq \sqrt{t})dt \\ &\overset{(4.13)}{=} \int_0^1 1 - t^{\frac{1}{2}2A(1/2)}dt \\ &= \frac{A(1/2)}{1 + A(1/2)}\end{aligned}$$

and since the first moment was calculated in (4.17), we know the variance

$$\begin{aligned}\mathrm{Var}\left(\max\left\{U, V\right\}\right) &= \mathbb{E}\big(\max\left\{U, V\right\}^2\big) - \mathbb{E}\big(\max\left\{U, V\right\}\big)^2 \\ &= \frac{A(1/2)}{1 + A(1/2)} - \left(\frac{2A(1/2)}{1 + 2A(1/2)}\right)^2 \\ &= \frac{A(1/2)(2A(1/2) - 1)^2 - 4\big[A(1/2)\big]^2(1 + A(1/2))}{(1 + A(1/2))(1 + 2A(1/2))^2}\end{aligned}$$

$$= \frac{4\big[A(1/2)\big]^3 + 4\big[A(1/2)\big]^2 + A(1/2) - 4\big[A(1/2)\big]^3 - 4\big[A(1/2)\big]^2}{(1+A(1/2))(1+2A(1/2))^2}$$
$$= \frac{A(1/2)}{(1+A(1/2))(1+2A(1/2))^2} < \infty.$$

Now the i.i.d. $\big\{ \max\big\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\big\}\big\}_{i=1,\ldots,n}$ is a sequence of random variables with the common expected value $\mathbb{E}\big(\max\big\{U,V\big\}\big)$ and finite variance $\sigma^2_{\max} := \operatorname{Var}\big(\max\big\{U,V\big\}\big)$. Thus (4.18) follows from the central limit theorem.

Next, we define the function

$$g : (0,1) \to \mathbb{R}, \quad t \mapsto 3 - (1-t)^{-1}.$$

Then

$$g\left(\frac{1}{n}\sum_{i=1}^{n} \max\big\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\big\}\right) = 3 - \left(1 - \frac{1}{n}\sum_{i=1}^{n} \max\big\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\big\}\right)^{-1} = \lambda_U^{\mathrm{FF}^*} \quad (4.19)$$

and

$$g\left(\mathbb{E}\big(\max\big\{U,V\big\}\big)\right) = 3 - \big(1 - \mathbb{E}\big(\max\big\{U,V\big\}\big)\big)^{-1} \overset{(4.15)}{=} 2 - \frac{2\,\mathbb{E}\big(\max\big\{U,V\big\}\big)}{2\left(1-\mathbb{E}\big(\max\big\{U,V\big\}\big)\right)}$$
$$\overset{(4.14)}{=} 2 - 2A(1/2) \overset{(3.39)}{=} \lambda_U. \quad (4.20)$$

Fix $t \in (0,1)$. Note that g is continuously differentiable on it's domain and its derivative at the point t is

$$g'(t) = \frac{1}{(1-t)^2}.$$

We write the first order approximation of the Taylor series of $g(t)$ at the point $\mathbb{E}\big(\max\big\{U,V\big\}\big)$. This is given by Taylor's theorem as

$$g(t) = g\left(\mathbb{E}\big(\max\big\{U,V\big\}\big)\right) + R_1(t) \overset{(4.20)}{=} \lambda_U + R_1(t),$$

where R_1 is the remainder term. Using the Lagrange form for the remainder, the above becomes

$$g(t) - \lambda_U = g'(\xi)\left(t - \mathbb{E}\big(\max\big\{U,V\big\}\big)\right),$$

where ξ is in the interval between t and $\mathbb{E}\big(\max\big\{U,V\big\}\big)$. By taking $t = \frac{1}{n}\sum_{i=1}^{n} \max\big\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\big\}$ and multiplying both sides of the equation by $\sqrt{n}$, we obtain

$$\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^{n} \max\big\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\big\}\right) - \lambda_U\right) \overset{(4.19)}{=} \sqrt{n}(\lambda_U^{\mathrm{FF}^*} - \lambda_U)$$
$$= g'(\xi)\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \max\big\{F_X(\widetilde{X}_i), F_Y(\widetilde{Y}_i)\big\} - \mathbb{E}\big(\max\big\{U,V\big\}\big)\right), \quad (4.21)$$

for every $n \in \mathbb{N}$.

Let us consider the behaviour of the sequence on the right-hand side. Recall that ξ lies in the interval between $\frac{1}{n}\sum_{i=1}^{n}\max\{F_X(\widetilde{X}_i),F_Y(\widetilde{Y}_i)\}$ and $\mathbb{E}(\max\{U,V\})$. This justifies the estimation

$$\left|\xi-\mathbb{E}(\max\{U,V\})\right|\leq\left|\frac{1}{n}\sum_{i=1}^{n}\max\{F_X(\widetilde{X}_i),F_Y(\widetilde{Y}_i)\}-\mathbb{E}(\max\{U,V\})\right|.$$

In the proof of Theorem 4.3.1 we applied the strong law of large numbers to show that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\max\{F_X(\widetilde{X}_i),F_Y(\widetilde{Y}_i)\}-\mathbb{E}(\max\{U,V\})\right|\xrightarrow[n\to\infty]{\text{a.s.}}0.$$

Combined with the above inequality, this implies $\xi\to\mathbb{E}(\max\{U,V\})$ almost surely, as $n\to\infty$. The continuity of g' implies that $g'(\xi)\to g'\left(\mathbb{E}(\max\{U,V\})\right)$ almost surely, as $n\to\infty$. Then by (4.18) and an application of Slutsky's theorem we obtain the convergence result

$$\begin{aligned}g'(\xi)\sqrt{n}&\left(\frac{1}{n}\sum_{i=1}^{n}\max\{F_X(\widetilde{X}_i),F_Y(\widetilde{Y}_i)\}-\mathbb{E}(\max\{U,V\})\right)\\&\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}\left(0,\left[g'\left(\mathbb{E}(\max\{U,V\})\right)\right]^2\sigma^2_{\max}\right).\end{aligned}$$

Name $\sigma^2:=\left[g'\left(\mathbb{E}(\max\{U,V\})\right)\right]^2\sigma^2_{\max}$. Now the equality of sequences (4.21) gives

$$\sqrt{n}(\lambda_U^{\mathrm{FF}^*}-\lambda_U)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}\left(0,\sigma^2\right).$$

We have all the needed information to calculate the variance explicitly. Since

$$\begin{aligned}\left[g'\left(\mathbb{E}(\max\{U,V\})\right)\right]^2&=\left(\frac{1}{\left(1-\mathbb{E}(\max\{U,V\})\right)^2}\right)^2\overset{(4.17)}{=}\left(\frac{1}{1-\frac{2A(1/2)}{1+2A(1/2)}}\right)^4\\&=(1+2A(1/2))^4,\end{aligned}$$

the variance becomes

$$\begin{aligned}\sigma^2&=\left[g'\left(\mathbb{E}(\max\{U,V\})\right)\right]^2\sigma^2_{\max}\\&=(1+2A(1/2))^4\,\frac{A(1/2)}{(1+A(1/2))(1+2A(1/2))^2}\\&=\frac{A(1/2)(1+2A(1/2))^2}{1+A(1/2)}.\end{aligned}$$

□

5 Simulation study

5.1 Plan of simulation study

We compare the performance of the estimators defined in Section 4 to each other. The sampling distributions are chosen so that the true value for the upper tail dependence coefficient can be solved with the knowledge of Section 3.3. The choice and parametrisation of the sampling distributions is explained in the following.

One scenario of interest for our investigation is that of tail independence. We sample three different tail independent distributions and observe the performance of our estimators within each distribution and across them. The samples from tail independent distributions under consideration will be denoted in the following way.

(i) Let In_n denote a sample of size n of a bivariate random vector with standard normal marginals and independence between the components. That is, In_n denotes a an i.i.d. sequence of random pairs of the form

$$(X_i, Y_i)_{i=1,\dots,n}, \quad \text{where } X_i, Y_i \sim \mathcal{N}(0,1) \text{ and } X_i \perp\!\!\!\perp Y_i,$$

for every $i = 1, \dots, n$. As shown by Lemma 3.4.1, the marginal distributions are arbitrary as far as the tail dependence coefficient is concerned and the important thing here is that the underlying copula is the Independence copula, which has been shown to be tail independent in Example 3.1.3.

(ii) Let N_n denote a sample of size n from the standard bivariate Gaussian distribution. That is, observations in sample N_n are an i.i.d. sequence

$$(X_i, Y_i)_{i=1,\dots,n} \sim \mathcal{N}_2(\mathbf{0}, \Sigma), \quad \text{where } \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This distribution was proven to exhibit tail independence in Example 3.1.5.

(iii) Let F_n denote a sample of size n from the frank copula with parameter $\theta = 1$. That is to say, sample F_n is an i.i.d sequence of random pairs

$$(X_i, Y_i)_{i=1,\dots,n}, \quad \text{where } F_{(X_i,Y_i)} = C_{\mathrm{Frank}}(U_i, V_i),\ U_i, V_i \sim \mathcal{U}(0,1),$$

for every $i = 1, \dots, n$. The frank copula was shown to exhibit tail independence in Example 3.4.2 for any parameter $\theta \in \mathbb{R} \setminus \{0\}$.

Another scenario of interest is the complete dependence, meaning the case where the marginals are almost surely the same. As shown in Example 3.3.2, this leads to the Comonotonic copula.

(iv) Let Co_n denote a sample of size n from the bivariate distribution defined by uniform marginals and the Comonotonic copula. In practice, we sample the exponential distribution with parameter 1. The observation vectors are then constructed from these samples so that both of their components are set equal to one observation from the exponential distribution. In this way, the sampling distribution has equal marginals which, according to 3.3.2, means their copula is the Comonotonic copula. Thus sample Co_n is an i.i.d sequence of random pairs

$$(X_i, Y_i)_{i=1,\dots,n}, \quad \text{where } F_{(X_i,Y_i)} = C_{\mathrm{Co}}(U_i, V_i),\ U_i, V_i \sim \mathcal{U}(0,1),$$

for every $i = 1, \dots, n$.

The third and final scenario of interest is how the estimators perform when $\lambda_U \in (0,1)$. The parametrisation of the following distributions has been decided on so that in both cases $\lambda_U = 0.5$ for ease of comparability.

(v) Let T_n represent a sample of size n from the centered bivariate t-distribution with $\nu = 1.5$ degrees of freedom, unit variances and correlation $\rho = 0.6045$. That is, the sample T_n is an i.i.d. sequence

$$(X_i, Y_i)_{i=1,\dots,n} \sim \mathbf{t}_\nu^n(\mu, \Sigma), \quad \text{where } \mu = (0,0) \text{ and } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The upper tail dependence coefficient of this distribution we have solved analytically in Example 3.4.3. By substituting the parameters ν and ρ into the formula arrived at in the example, we get $\lambda_U = 0.5000034\ldots \approx 0.5$.

(vi) Let Gum_n represent a sample of size n from the 2-dimensional Gumbel copula with paramater $\theta = (\log_2(3/2))^{-1} \approx 1.7095$. So the sample Gum_n is an i.i.d sequence of random pairs

$$(X_i, Y_i)_{i=1,\dots,n}, \quad \text{where } F_{(X,Y)} = C_{\text{Gumbel}}(U_i, V_i),\ U_i, V_i \sim \mathcal{U}(0,1),$$

for every $i = 1, \dots, n$. The tail dependence coefficient of the Gumbel copula was solved analytically in Example 3.4.4. By substituting the parameter θ into the formula given in the example we get $\lambda_U = 0.5$.

We now have three tail independent sampling distributions and three tail dependent sampling distributions. This should let us evaluate estimator performance over a wide variety of samples. The distributions can also be categorised into ones with extreme value copulas and ones with non-extreme value copulas. We proved that the Independence copula, Comonotonic copula, and the Gumbel copula are extreme value copulas in Example 3.3.8.

In Section 4 we discussed the asymptotic properties of the FF estimator and mentioned those of the CFG estimator. These properties are proven under the assumption that the underlying copula is an extreme value copula. Thus this asymptotic behaviour should particularly be expressed when calculating the CFG and FF estimators for varying sample sizes of In_n, Co_n and Gum_n. In contrast, the sec and log estimators should portray strong consistency, that is, convergence toward the analytically solved value of the tail dependence coefficient, when applied to any of the samples (and varying the sample size.)

We generate 1000 samples from each sampling distribution at three different sample sizes: $n = 250, n = 1000$ and $n = 5000$. We then use each of these 1000 samples to calculate 1000 of each estimator at every sample size. We then compare the following values for each estimator and at each sample size. We determine the mean of estimates which were calculated from the 1000 samples. Denote 1000 estimator values by $\hat{\lambda}_i$, where $i = 1, \dots, 1000$. Then

$$\mu\left(\hat{\lambda}_U\right) := \frac{1}{1000} \sum_{i=1}^{1000} \hat{\lambda}_i.$$

Using the mean, we calculate the difference of the true value of λ_U and the mean of estimated values i.e. the *bias*:

$$\text{BIAS}\left(\hat{\lambda}_U\right) := \mu\left(\hat{\lambda}_U\right) - \lambda_U,$$

We also calculate the sample standard deviation of the 1000 estimates

$$\hat{\sigma}\left(\hat{\lambda}_U\right) := \sqrt{\frac{1}{n-1} \sum_{i=1}^{1000} \left(\hat{\lambda}_{n,i} - \mu\left(\hat{\lambda}_U\right)\right)^2}.$$

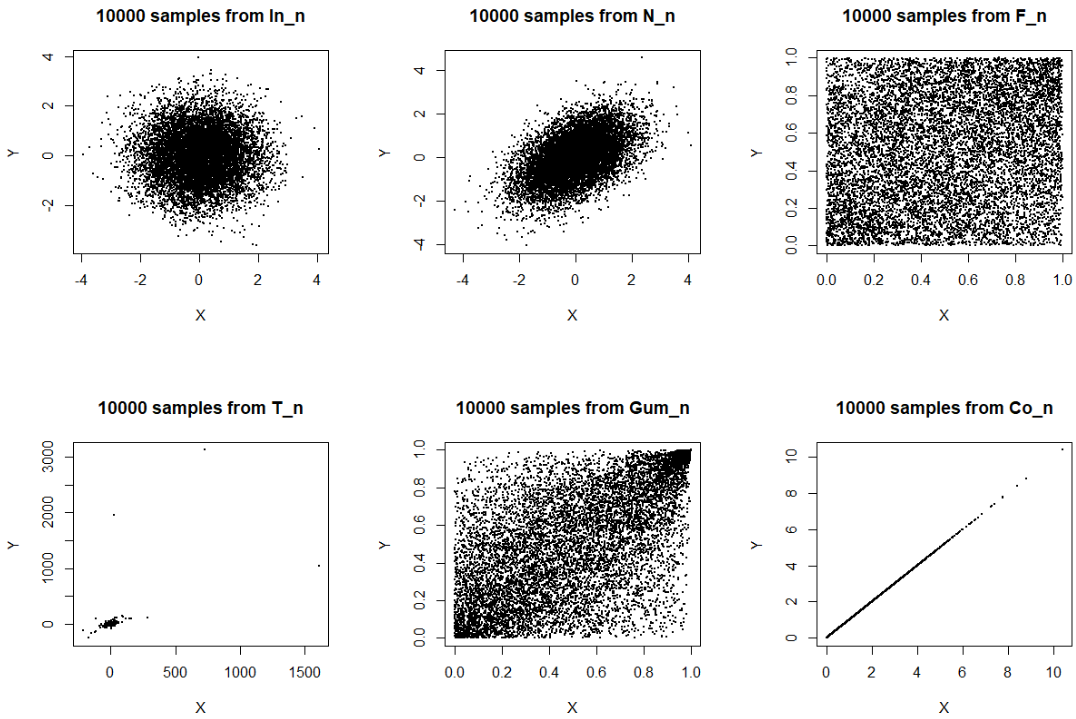


Figure 5.1: 10000 observations of each sample, with the X-coordinate plotted against Y-coordinate.

Multiple estimator values are calculated to allow us to investigate the behaviour of the estimators for a relatively large variety of data. As we see in the results of table 5.1, the standard deviations can be considerably large. This means that an estimate calculated from one sample of size 1000 may differ considerably from an estimate calculated from another sample of size 1000, even when the sampling distribution stays unchanged.

The standard deviations we calculate let us consider how likely it is to end up with an estimate close to the calculated mean of the estimator. Equivalently, it lets us know how likely it is for an estimator value to be realised such that its bias is close to the calculated bias of the estimator.

We compare the estimators to each other using the bias and sample standard deviation. The mean is left out of consideration. This is because the actual estimates are not as interesting to us as the information of how close the estimates are to the true parameter value. The latter information is conveyed by the bias. The bias and the true value of the parameter can be added together to construct the mean, if one is interested in it. In general, the smaller the bias and sample standard deviation are, the better the estimator is interpreted to have performed.

5.1.1 Choice of the number of block maxima

The method of block maxima was introduced in Section 4.1. One can justify applying estimation methods, which are based on the assumption of extreme value data, to non-extreme value data by transforming the given sample through the method of block maxima. In the case of this simulation study, the estimation methods based on the assumption of extreme value data are the CFG and FF estimators, while the non-extreme value samples they are applied to are N_n, F_n and T_n, for $n = 250, 1000, 5000$. Thus a decision on the number of block maxima to be taken needs to be made.

The number of block maxima used in the simulation study is 250 and it is kept constant across all distributions and sample sizes. Therefore one can hypothesize that the results of the CFG and FF estimators, when applied to non-extreme value sampling distributions, can be improved with more intelligent and scenario-specific choices of the number of block maxima. Most egregiously, at the smallest sample size used in the study, $n = 250$, this causes the block maxima to be effectively not taken at all. This potentially negatively affects the performance of the CFG and FF estimators, as it represents precisely the kind of case where they are applied to unaltered non-extreme value data.

This decision of the number of block maxima is made because to investigate optimal choices of block maxima is outside the subject matter of this thesis and this specific number of block maxima seems to perform well. This is to say, that it considerably lessens the bias of the CFG and FF estimators when applied to the non-extreme value sampling distributions N_n, F_n and T_n, while keeping the standard deviation reasonably small. Additionally, one could argue, that the difficulty of applying the method of block maxima at small sample sizes is inevitably a hindrance for the applicability of the CFG and FF estimators (when used on non-extreme value data). Thus their non-optimal performance under small samples here could be viewed as intrinsic to their poor performance even in the case where the method of block maxima is applied more intelligently.

5.2 Results of simulation study and their interpretation

We interpret the results of the simulation study, which are presented in Table 5.1.

Overall, the sec estimator performed the worst of the estimators across all distributions with respect to both bias and standard deviation. It did beat the CFG and FF estimators in bias when applied to the non-extreme value sampling distributions N_n, F_n and T_n at small sample sizes ($n = 250$). Even so, it has considerably larger or equivalent standard deviation in these scenarios than the other estimators and when the sample size is increased, the other estimators perform better in both benchmarks. Moreover, the superiority of the sec estimator under small sample sizes could be explained by the suboptimal performance of the CFG and FF estimators caused by the choice of block maxima discussed in Section 5.1.

The log estimator consistently outperformed the sec estimator across all sampling distributions. It beat the CFG and FF estimators, particularly when applied to non-extreme value data. It is especially competitive in bias at small sample sizes, but it is exactly in these cases that the log estimator portrays relatively large standard deviation. Consequently, the log estimator seems unreliable at small sample sizes.

The CFG estimator performed best on the extreme value data of samples In_n and Gum_n. When applied to these samples, it beats the sec and log estimators in both bias and standard deviation, particularly at larger sample sizes.

The FF estimator performs best on the extreme value data of samples In_n and Gum_n. What is particularly impressive in these cases, is its performance in bias under small sample sizes. At sample size $n = 250$, its bias is already under 1% (see Figure 5.2). Its sample standard deviation across all samples, however, is consistently equivalent or worse than that of the CFG estimator's.

			sec		log		CFG		FF	
Dist.	λ_U	n	BIAS	$\hat{\sigma}$	BIAS	$\hat{\sigma}$	BIAS	$\hat{\sigma}$	BIAS	$\hat{\sigma}$
In_n	0	250	0.0831	0.0758	0.0183	0.0632	0.0298	0.0484	0.0080	0.0596
		1000	0.0527	0.0472	0.0092	0.0320	0.0080	0.0243	0.0004	0.0298
		5000	0.0350	0.0349	0.0038	0.0139	0.0015	0.0109	-0.0003	0.0133
N_n	0	250	0.3023	0.1078	0.2506	0.1081	0.3810	0.0408	0.4063	0.0421
		1000	0.2545	0.0875	0.2072	0.0673	0.2708	0.0452	0.2881	0.0504
		5000	0.2396	0.1049	0.1901	0.0528	0.1717	0.0462	0.1754	0.0529
F_n	0	250	0.1164	0.0903	0.0452	0.0758	0.1446	0.0480	0.1525	0.0558
		1000	0.0885	0.0673	0.0296	0.0418	0.0803	0.0490	0.0764	0.0598
		5000	0.0593	0.0623	0.0175	0.0215	0.0412	0.0470	0.0247	0.0590
T_n	0.5	250	0.0248	0.0997	0.0027	0.1060	0.0260	0.0434	0.0254	0.0428
		1000	0.0163	0.0728	0.0037	0.0742	-0.0133	0.0416	-0.0238	0.0425
		5000	0.0097	0.0416	0.0018	0.0422	-0.0044	0.0390	-0.0106	0.0396
Gum_n	0.5	250	0.0408	0.1041	0.0213	0.1090	0.0082	0.0374	0.0010	0.0388
		1000	0.0222	0.0718	0.0110	0.0726	0.0024	0.0188	0.0004	0.0192
		5000	0.0167	0.0406	0.0091	0.0409	0.0002	0.0084	-0.0001	0.0085

Table 5.1: The bias and sample standard deviation of tail dependence estimators calculated from different distributions at three different sample sizes each. The true value of the (upper) tail dependence coefficient of each distribution is listed in the second column. All entries for samples from the distribution Co_n were equal to zero and are therefore omitted. The smaller the bias and standard deviation, the better the estimator is interpreted to have performed.

5.2.1 Paired behaviour

Even though the log estimator performs consistently better than the sec estimator, the rates of convergence of their biases and standard deviations over the chosen sample sizes n seem similar. The same could be said for the CFG and FF estimators. This paired behaviour can be observed in Figures 5.4 and 5.5.

The log estimator seems to beat the CFG and FF estimators when applied to the non-extreme value samples F_n and T_n. Generally, the CFG and FF estimators performed better than the log (and sec) estimator over the extreme value samples In_n and Gum_n. This dichotomy is perhaps to be expected. After all, the sec and log estimators are both based on the same idea of approximating the derivative of the diagonal of the copula with a suitable transformation of the empirical copula. Likewise, the CFG and FF estimators were both specifically motivated by the assumption of an extreme value sampling distribution.

5.3 Further investigation through simulation studies

5.3.1 The effect of the level of tail dependence on estimator accuracy

There are some previously unstated properties of the estimators which come to light through further investigation of Table 5.1. One of them is the ability of all estimators to identify comonotonicity at all sample sizes.

The results for samples taken from the distribution Co_n are not included in Table 5.1, since all entries were equal to zero. This gives a positive assessment of the accuracy of the estimators in the case of full dependence.

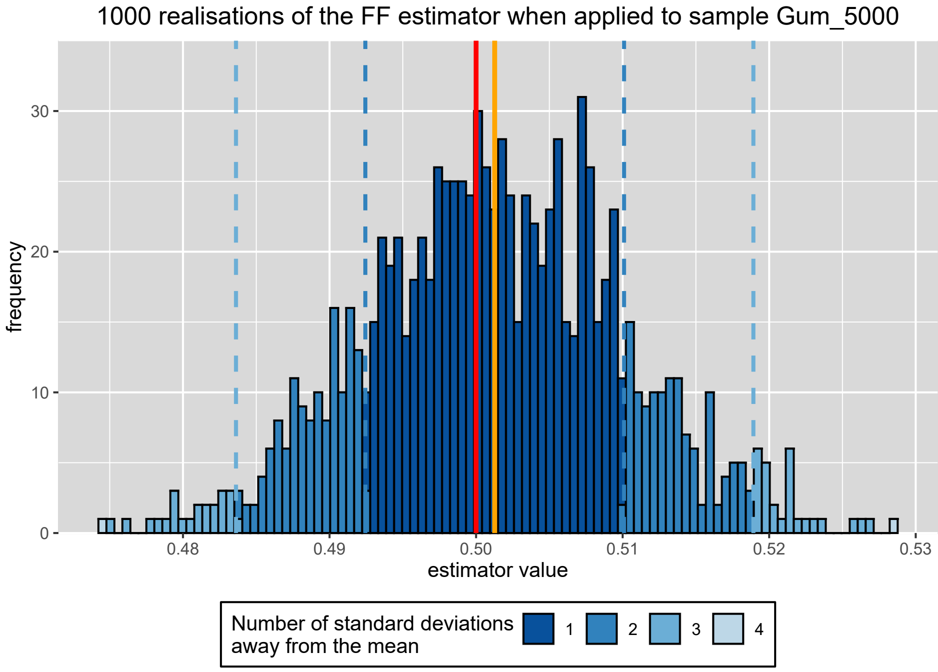


Figure 5.2: Histogram depicting an example of good estimator performance. The mean of estimator values is drawn in red and the true value of the tail dependence coefficient is drawn in orange.

We can reason that this occurs because for a comonotonic sample $(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\dots,n}$, with continuous marginals F_X and F_Y, the components are almost surely equal. This has implications for the behaviour of the empirical copula. It implies almost sure equality for the ranks of the observations:

$$R_X^{(i)} := \sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq \widetilde{X}_i\} \overset{\text{a.s}}{=} \sum_{j=1}^{n} \mathbb{1}\{\widetilde{Y}_j \leq \widetilde{Y}_i\} =: R_Y^{(i)}. \tag{5.1}$$

Furthermore, the sampling is done from a distribution with continuous marginals which means that the probability of two observations having the same realised value is zero, i.e. $\mathbb{P}(\widetilde{X}_i = \widetilde{X}_j) = 0$, for all $i, j = 1, \dots, n$, for which $i \neq j$. This means that the ranked observations are almost surely the set of numbers from 1 to n:

$$\{R_X^{(i)} |\, i = 1, \dots, n\} = \left\{ \sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq \widetilde{X}_i\} \middle|\, i = 1, \dots, n \right\} \overset{\text{a.s.}}{=} \{1, \dots, n\}. \tag{5.2}$$

By simply plugging this information into the definition of each estimator, we notice the following

(i) sec: Let $k \in \{1, \dots, n-1\}$. The definition of the empirical copula combined with the above, yields

$$\hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right) \overset{(4.1)}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{\frac{R_X^{(i)}}{n} \leq \frac{n-k}{n}, \frac{R_Y^{(i)}}{n} \leq \frac{n-k}{n}\right\}$$

$$\overset{\text{a.s.}}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{R_X^{(i)} \leq n-k\right\}$$
$$\overset{\text{a.s}}{=} \frac{n-k}{n},$$

where the last equality holds because the number $n-k$ is in the set of numbers $\{1,\ldots,n\}$ and so by (5.2) there are exactly $n-k$ ranks $R_X^{(i)}$ which are smaller or equal to $n-k$. As a consequence of this, the diagonal of the empirical copula constructed from the comonotonic sample is an identity function, at least on the set {1,. . . ,n}, for any sample size. Thus

$$\hat{\lambda}_U^{\text{sec}} \overset{(4.5)}{=} 2 - \frac{1-\hat{C}_n\left(\frac{n-k}{n},\frac{n-k}{n}\right)}{1-\frac{n-k}{n}} \overset{a.s.}{=} 2-1=1.$$

(ii) log: Using the same reasoning as above, we can easily see that for any sample size n and choice of threshold $k \in \{1,\ldots,n-1\}$,

$$\hat{\lambda}_U^{\log} \overset{(4.6)}{=} 2 - \frac{\log\left(\hat{C}_n(\frac{n-k}{n},\frac{n-k}{n})\right)}{\log\left(\frac{n-k}{n}\right)} \overset{\text{a.s.}}{=} 2-1=1.$$

(iii) CFG: According to (5.1), for comonotonic samples

$$\hat{U}_i = \frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{X}_j \leq \widetilde{X}_i\} \overset{\text{a.s.}}{=} \frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{Y}_j \leq \widetilde{Y}_i\} = \hat{V}_i,$$

for every $i = 1\ldots,n$. Thus,

$$\lambda_U^{\text{CFG}} \overset{(4.12)}{=} 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n}\log\left(\sqrt{\log\frac{1}{\hat{U}_i}\log\frac{1}{\hat{V}_i}} \Big/ \log\frac{1}{\max\left\{\hat{U}_i,\hat{V}_i\right\}^2}\right)\right\}$$
$$\overset{\text{a.s.}}{=} 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n}\log\left(\left|\log\frac{1}{\hat{U}_i}\right| \Big/ 2\log\frac{1}{\hat{U}_i}\right)\right\}$$
$$= 2 - 2\exp\left\{\frac{1}{n} n \log\left(\frac{1}{2}\right)\right\} = 2-1=1.$$

(iv) FF: Similar to the above, we know that $\hat{U}_i \overset{\text{a.s.}}{=} \hat{V}_i$. These were defined as

$$\hat{U}_i = \frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{X_j} \leq \widetilde{X_i}\} \overset{\text{a.s.}}{=} \hat{V}_i.$$

We observe that for an even sample size n,

$$\frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}(1+2+3+\ldots+(n/2)+(n-((n/2)-1))+\ldots+(n-2)+(n-1)+n)$$
$$= \frac{1}{n}((n+1)+((n-1)+2)+((n-2)+3)+\ldots+((n-((n/2)-1))+(n/2)))$$
$$= \frac{1}{n}\left(\frac{n}{2}(n+1)\right) = \frac{n+1}{2}.$$

(The sample sizes used in our study are even. This formula does also apply in the odd case. The proof is similar.) By making use of this trick, we calculate

$$
\begin{aligned}
\lambda_U^{\mathrm{FF}} &\overset{(4.15)}{=} 3 - \left(1 - \frac{1}{n}\sum_{i=1}^{n} \max\left\{\hat{U}_i, \hat{V}_i\right\}\right)^{-1} \overset{\text{a.s.}}{=} 3 - \left(1 - \sum_{i=1}^{n} \frac{1}{n}\hat{U}_i\right)^{-1} \\
&= 3 - \left(1 - \sum_{i=1}^{n} \frac{1}{n}\left[\frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}\{\widetilde{X_j} \leq \widetilde{X_i}\}\right]\right)^{-1} \\
&\overset{(5.2)}{=} 3 - \left(1 - \frac{1}{n+1}\left[\frac{1}{n}\sum_{i=1}^{n} i\right]\right)^{-1} \\
&= 3 - \left(1 - \frac{1}{n+1}\left[\frac{n+1}{2}\right]\right)^{-1} = 3 - 2 = 1.
\end{aligned}
$$

Therefore it appears that every estimator identifies comonotonicity at any sample size. Comonotonicity is, however, a quite academic example of a dependence structure. It is easy to imagine that the comonotonicity of a given data set will often, if not always, be noticed before any tail dependence estimation is attempted. Therefore these estimators are more interesting in the in-between of independence and complete dependence. Nevertheless, estimator performance in these extreme cases does give us a piece of the picture of how these estimators behave at different dependence levels.

By far the worst performance of the estimators was brought out by the samples N_n, with $n = 250, 1000, 5000$. Even at the largest sample size, all estimators were greater than the real tail dependence coefficient value by at least 17%. The sample standard deviations are such that at least 95% of the realised estimator values belong to an interval above the analytical value of the tail dependence coefficient (see Figure 5.3). These results do not significantly improve if the sample size is increased up to $n = 10000$.

In [9], Frahm, Junker and Schmidt apply the sec, log and CFG estimators to a tail independent elliptical distribution which they refer to as asymmetric generalised hyperbolic distribution. The correlation parameter is set, similar to our Gaussian distribution, to $\rho = 0.5$. Their results are similar to our results for N_n. This suggests that there is something about elliptically distributed data that makes its tail dependence, or lack thereof, particularly elusive for these estimators.

It would seem that estimator performance is poor, not only on elliptically distributed data but tail independent samples in general. The log, CFG and FF estimators performed fairly well when applied to In_{5000}. At smaller sample sizes, however, it seems very possible to end up with values for even these estimators which would indicate there to be some tail dependence. This is further supported by the results of estimation done on the samples F_n, for whom the only reasonably good estimation results are achieved by the log estimator at sample size $n = 5000$.

The estimation results tend to err on the side of tail dependence. This is not only seen as excellent performance on the comonotonic sample Co_n. It is also seen in the estimation results for the T_n and Gum_n distributions (see, for example, the Figure 5.2). The estimators perform comparatively well when applied to these samples as opposed to the tail independent ones. The reverse seems to not occur, in that, none of the results show tail independence being indicated where it's not truly present.

Let's investigate this phenomenon further: We sample i.i.d. observations from four Gumbel copulas with different parameters $\theta \geq 1$. The parameters are chosen so that the four tail

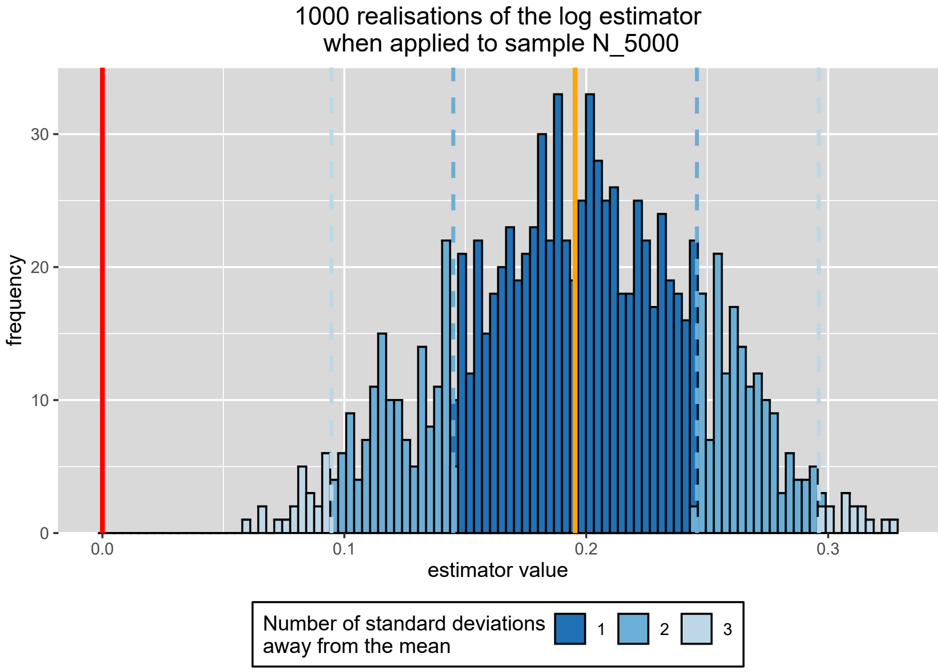


Figure 5.3: Histogram depicting an example of poor estimator performance. Notice how the true tail dependence coefficient value (drawn in red) is more than three sample standard deviations away from the mean of the estimates (drawn in orange). This results in at least 98.7% of the estimator values lying inside an interval entirely above the parameter value.

dependence coefficients of the sampling distributions are 1/4, 1/2, 3/4 and ≈ 1. We sample each distribution 1000 times at the sample size $n = 1000$. For every sample, we calculate the value of our estimators and compare the bias and standard deviation against the tail dependence coefficient values. The results are plotted in Figure 5.4.
An equivalent procedure performed on multivariate t-distributed sampling distributions produces similar results. Thus it would seem that our estimators do in fact perform better the more tail dependent data they are applied to. This is also supported by the show of poor performance on tail independent samples $\mathrm{In}_n, \mathrm{N}_n, \mathrm{F}_n$.

Consequently, when investigating the tail dependence structure of a given data set, it makes sense to first test for independence in the data. If the observation components are deemed independent, calculating estimators for the tail dependence coefficients becomes unnecessary. This way confusion caused by large estimator values in the case of independence between observation components can be mitigated.

This also means that additional statistical tests for tail independence itself are necessary and should be applied to data before tail dependence coefficient estimation. Through such additional testing, one can avoid falsely assigning tail dependence to distributions which, analytically speaking, do not exhibit it.

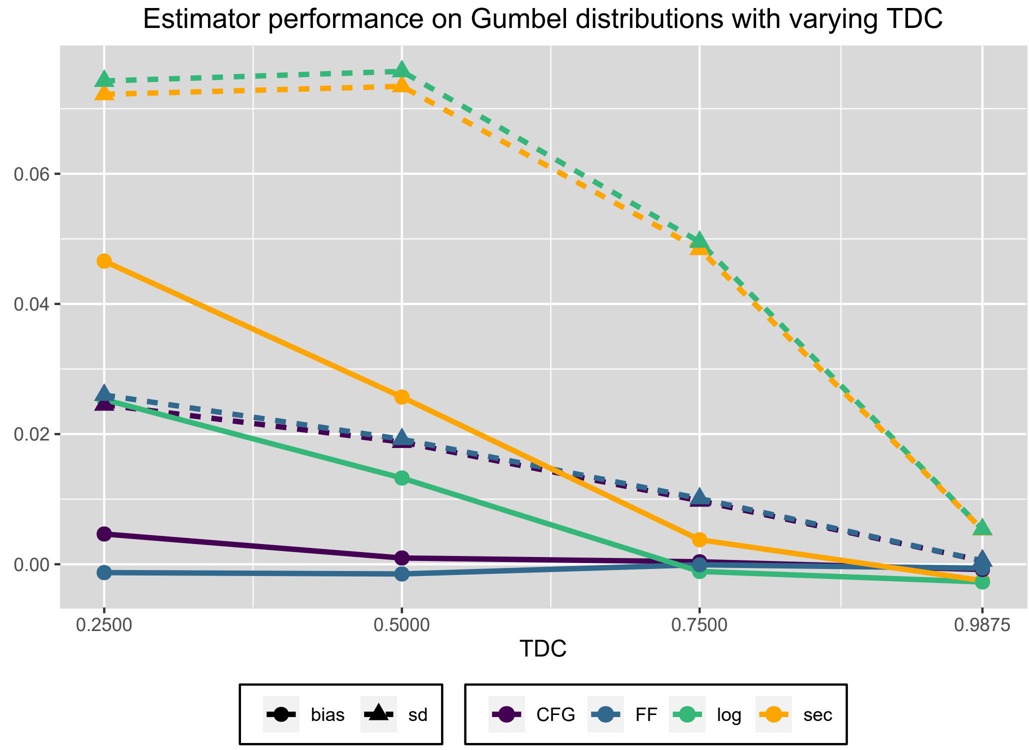


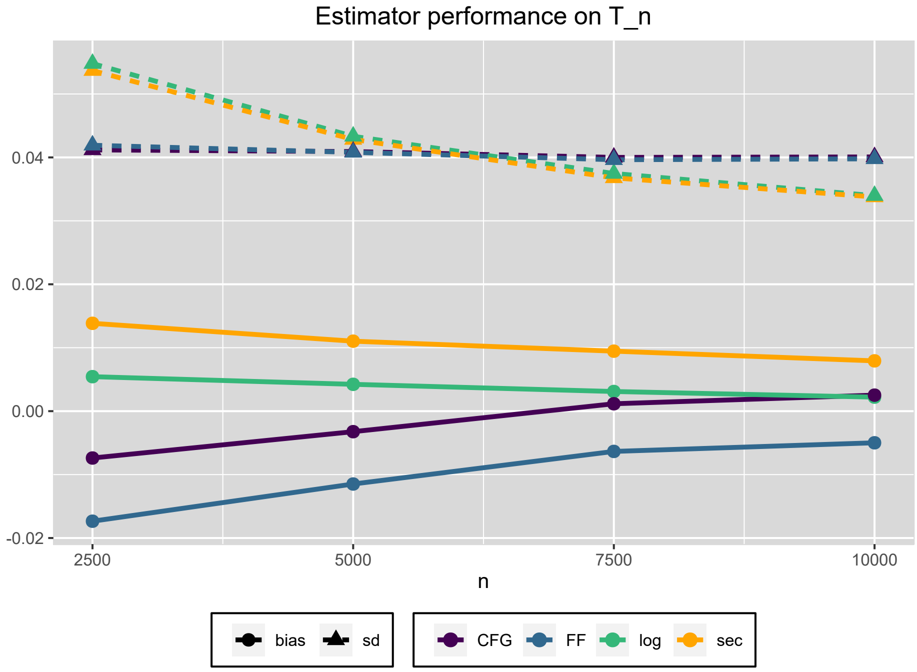
Figure 5.4: The BIAS and $\hat{\sigma}$ of estimators plotted against the true value of the tail dependence coefficient. As the tail dependence coefficient grows, the estimators perform better. This is seen as convergence toward zero for both the bias and standard deviation of all estimators.

5.3.2 Further investigation into the asymptotic properties of the estimators

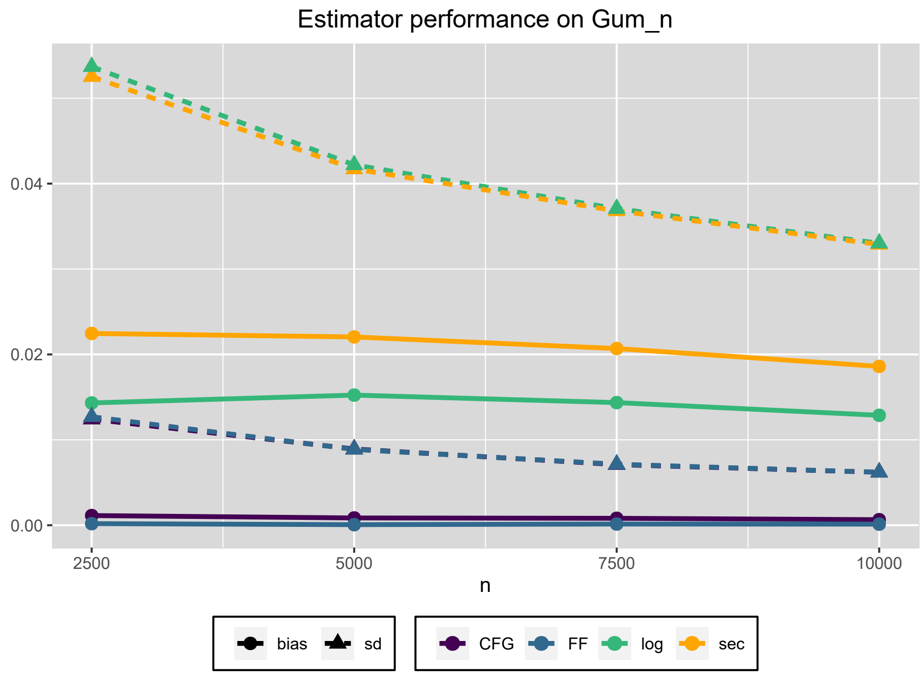
The results of Table 5.1 indicate that the estimators posses the desired asymptotic properties discussed in Section 4.3. Namely, both the bias and the sample standard deviation decrease as the sample size increases.

We now investigate further the asymptotic properties of the estimators by generating 1000 samples from two distributions at four different sample sizes $n = 2500, 5000, 7500$ and 10000. Estimates and their biases and sample standard deviations are then calculated. The sampling distributions used are T_n and Gum_n. They represent a non-extreme value sampling distribution and an extreme value distribution respectively. The parametrisation is kept the same as it was in Section 5.1 so that the tail dependence coefficient for both sampling distributions is 0.5. The results are plotted in Figure 5.5.

There is a predictable dichotomy in the performance of the estimators on the two sampling distributions. The superior performance of sec and log on the non-extreme value sampling distribution T_n and conversely the better performance of CFG and FF on the extreme value distribution Gum_n can be observed in the order of their biases and standard deviations in each plot.



(a) Asymptotic behaviour of estimators when applied to the multivariate t-distributed samples T_n.



(b) Asymptotic behaviour of estimators when applied to the Gumbel copula-distributed samples Gum_n

Figure 5.5: Estimator bias and sample standard deviation plotted against sample size.

And indeed, the asymptotic properties of the estimators can also be observed. This can be seen as the convergence toward zero of the biases of all estimators, which corresponds to the means of the estimators converging toward the true parameter value. The standard deviations converge toward zero as well, only at a slower rate.

The exception is the behaviour of the CFG and FF estimators in Plot (a). The sample standard deviations do not converge toward zero but instead stay approximately constant. This could be the effect of a poor choice of the number of block maxima. It could be further exacerbated by the heavy tails of the sampling distribution T_n, which are evident from the scatter plot of Figure 5.1. Repetitions of simulation and estimation confirm that indeed the samples T_n produce quite a variety of estimation results. Either way, the sec and log estimators consistently beat the CFG and FF estimators in accuracy and thus this phenomenon speaks once more for the superiority of the sec and log estimators in the case of a non-extreme value sampling distribution.

6 Conclusions

Based on the simulation study of Section 5, the tail dependence coefficient estimators introduced in Section 4 are useful in practice under certain circumstances. By comparing the results of the simulation study (Table 5.1) it can be seen that the log estimator $\lambda_U^{\log}$ performs well in the presence of tail dependence. There seems to be no reason to use the sec estimator instead of the log estimator as the latter consistently out-performs the former. The estimators λ_U^{CFG} and λ_U^{FF} work wonderfully with extreme value data (see, for example, Figure 5.2).

However, it is evident that the tail dependence coefficient estimators by no means offer exhaustive answers to questions about tail dependence. One primary concern is that one needs to be working with large sample sizes to get reliable estimates. In [9], Frahm, Junker and Schmidt recommend using more distribution-specific estimation methods in the face of small sample sizes. In practice, this would mean identifying something about the distribution of the observations by using statistical tests and then utilising the acquired knowledge of the common distribution to estimate distribution-specific parameters which characterise the dependence structure.

Another difficulty associated with the presented estimators is that, in a non-extreme value framework, the user must select a suitable threshold or number of block maxima parameter (or perhaps a threshold for the peak-over-threshold method).

The worst cases of inaccuracy in our simulation study are caused by tail independent samples. It seems that the less tail dependent a data set is the worse the estimators perform. This same phenomenon was also encountered by the authors of [9] and also Ferreira in [8]. The proposed solution of all authors was to test for tail independence separately of tail dependence estimation.

Appendices

Appendix A Notation

Range(f)	The range of function $f : A \to B$, i.e. the set $\{b \in B \mid \exists a \in A, f(a) = b\}$.
$\overset{\mathcal{D}}{=}$	Equality in distribution.
$\sim$	For example, $X \sim \mathcal{N}(0, 1)$, meaning X has distribution $\mathcal{N}(0, 1)$.
$\sqcup$	Disjoint union. For example, $A \sqcup B$ is the union of A and B, where $A \cap B = \emptyset$.
$\perp\!\!\!\perp$	Denotes independence. For example, $X \perp\!\!\!\perp Y$, meaning that X and Y are independent.
$\hat{}$	Used on symbols denoting estimators. For example, $\hat{\lambda}$ is an estimator of λ.
$\mathbb{1}$	Indicator function.
F	A distribution function.
$\overline{F}$	Tail distribution function, that is, $\overline{F} = 1 - F$.
F^{-1}	Inverse function of F, that is, the function for which $F \circ F^{-1} = F \circ F^{-1} = \text{id}$.
$F^{\leftarrow}$	Generalised inverse function of F. Defined in 2.1.1.
Φ	The distribution function of $\mathcal{N}(0, 1)$.
ϕ	The density function of $\mathcal{N}(0, 1)$.
C	Copula, defined in 3.2.1.
$\widetilde{C}$	Survival copula, defined in 3.2.4.
$\mathcal{U}(a, b)$	The (univariate) uniform distribution over the interval $(a, b) \subset \mathbb{R}$.
$\mathcal{N}(\mu, \sigma^2)$	The (univariate) normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$.
$\mathcal{N}^d(\mu, \Sigma)$	The d-dimensional normal distribution with mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ.
Exp(α)	The exponential distribution with parameter $\alpha > 0$.
$(\widetilde{X}_i, \widetilde{Y}_i)_{i=1,\ldots,n}$	A sample of $n \in \mathbb{N}$ observations of the form $(\widetilde{X}_i, \widetilde{Y}_i)$, where $i = 1, \ldots, n$.
$\hat{F}_n$	Empirical distribution function calculated from a sample of size n. Defined in 2.3.1.
$\hat{C}_n$	Empirical copula calculated from a sample of size n, defined in (4.1).
λ_U	Upper tail dependence coefficient, defined in 3.1.2.
λ_L	Lower tail dependence coefficient, defined in 3.1.2.
λ_U^{sec}	The sec estimator, defined in (4.5).
$\lambda_U^{\log}$	The log estimator, defined in (4.6).
λ_U^{CFG}	The CFG estimator, defined in (4.12).
λ_U^{FF}	The FF estimator, defined in (4.15).

Appendix B R-code

```
#R version 3.5.0 (2018-04-23)
#Platform: x86_64-w64-mingw32/x64 (64-bit)
#Running under: Windows >= 8 x64 (build 9200)
#package versions used:
# ggplot2_3.3.3
# RColorBrewer_1.1-2
# viridis_0.6.1
# viridisLite_0.4.0
# xtable_1.8-4
# mvtnorm_1.1-1
# copula_1.0-1

#Load packages-------------------------------------------------------------------
require("copula") # for Frank and Gumbel copulas
require("mvtnorm") # for mvnorm and mvt functions
require("xtable") # for exporting tables into LaTeX
require("viridis") # for color palettes used in ggplotting
require("RColorBrewer") # same as above
require("ggplot2") # for plotting

#Miscellaneous functions---------------------------------------------------------

#alarm sound for running slow code
beep <- function(n = 3){
  for(i in seq(n)){
    system("rundll32 user32.dll,MessageBeep -1")
  }
  Sys.sleep(.5)
}

#calculate m block maxima from input vector
blockMaxVec <- function(m,rhsVec){
  l <- length(rhsVec)/m
  if (l%%1!=0){return(NULL)}
  else{
    returnVec <- max(rhsVec[1:l])
    if(m > 1){
      for (i in 2:m){
        returnVec <- c(returnVec,max(rhsVec[((i-1)*l +1):(i*l)]))
      }
    }
    return(returnVec)
  }
}

#calculate m component-wise block maxima from the given array (ncol has to be 2)
```

```r
blockMaxAr <- function(m,rhsVec){
  return(cbind(blockMaxVec(m,rhsVec[,1]),blockMaxVec(m,rhsVec[,2])))
}

#Estimators-----------------------------------------------------------------

#secant estimator evaluated at given threshold
secEstAsFunOfThreshold <- function(obsMat, nBlocks=0, thresholds){
  #checking input parameters
  if(nBlocks==0){nBlocks = nrow(obsMat)}
  if(sum((thresholds>nrow(obsMat)),na.rm=TRUE) > 0){return(NULL)}

  #evaluation of copula at relevant point
  evalPoints <- (nBlocks-thresholds)/nBlocks
  copulaVals <- C.n(u=cbind(evalPoints,evalPoints),
                  X=pobs(blockMaxAr(m=nBlocks,rhsVec=obsMat)))#!CHANGED!
  return(2 - ((1- copulaVals)/(1-evalPoints)))
}

#set the required minimun number of blocks for the threshold algorithm to work
constMinBlocks <- 100

#sec
TDCsecEst <- function(obsMat, nBlocks=0){
  nBlocks<-0
  #by default the blockmaxima will have no effetc on the result
  if(nBlocks==0){nBlocks <- nrow(obsMat)}
  #seems the sample size has to be atleast 100 for this function to work
  if(nBlocks < constMinBlocks){
    print("Sample size too small or too few blocks taken!")
    return(NULL)
  }

  #we calculate the estimate as an average of estimates in the smoothing
  # bandwith around the desired threshold
  #set constants
  sampleSize<-nBlocks #this has to be due to the definition of evalPoint
  smootheningBandwidth <- floor(0.005*sampleSize)
  #makeshift trick:
  if(smootheningBandwidth<1){smootheningBandwidth <- 1}
  plateauLength <- floor(sqrt(sampleSize-2*smootheningBandwidth))

  #calculate estimates for sec for every possible threshold k (I've modified
  # this so that we only go through 1/4th of the possible thresholds. If the
  # plateau is found past this it would stand to reason that it would no
  # longer be relevant)
  estimates <- secEstAsFunOfThreshold(obsMat, sampleSize, (1:(sampleSize/4)))

  #calculate smoothing averages an their standard deviation
```

```
  estAverages <- rep(x=0,times=(sampleSize/4)-2*smootheningBandwidth)
  for(k in (1:length(estAverages))){
    estAverages[k] <- mean(estimates[k:(k+2*smootheningBandwidth)])
  }

  standDevOfAvgs <- sd(estAverages)

  #test the different thresholds as per the criteria of Frahm et al.
  chosenThreshold <- 0
  for(k in (1:(length(estAverages) - plateauLength + 1)) ){
    if(sum(estAverages[(k+1):(k+plateauLength-1)] - estAverages[k]) <=
       2*standDevOfAvgs){
      chosenThreshold <- k
      break
    }
  }

  if(chosenThreshold==0){return(0)}
  else{
    return(data.frame(
      TDCsecEst =
        mean(estAverages[chosenThreshold:(chosenThreshold+plateauLength-1)]),
      chosenThreshold,
      plateauLength,
      sampleSize,
      smootheningBandwidth))
  }
}

#log
#logarithm estimator evaluated at given threshold
logEstAsFunOfThreshold <- function(obsMat, nBlocks=0, thresholds){
  #checking input parameters
  if(nBlocks==0){nBlocks = nrow(obsMat)}
  if(sum((thresholds>nrow(obsMat)),na.rm=TRUE) > 0){return(NULL)}

  #evaluation of copula at relevant point
  evalPoints <- (nBlocks-thresholds)/nBlocks
  copulaVals <- C.n(u=cbind(evalPoints,evalPoints),
                  X=pobs(blockMaxAr(m=nBlocks,rhsVec=obsMat)))#copula pkg
  return(2 - (log(copulaVals)/log(evalPoints)))
}

#This is just boiler plate of the sec equivalent above
TDClogEst <- function(obsMat, nBlocks=0){
  nBlocks<-0
  #by default the blockmaxima will have no effetc on the result
  if(nBlocks==0){nBlocks <- nrow(obsMat)}
  #seems the sample size has to be atleast 100 for this function to work
```

```r
  if(nBlocks < constMinBlocks){
    print("Sample size too small or too few blocks taken!")
    return(NULL)
  }

  #we calculate the estimate as an average of estimates in the smoothing
  # bandwith around the desired threshold
  #set constants
  sampleSize<-nBlocks #this has to be due to the definition of evalPoint
  smootheningBandwidth <- floor(0.005*sampleSize)
  #makeshift trick:
  if(smootheningBandwidth<1){smootheningBandwidth <- 1}
  plateauLength <- floor(sqrt(sampleSize-2*smootheningBandwidth))

  estimates <- logEstAsFunOfThreshold(obsMat, sampleSize, (1:(sampleSize/4)))

  #calculate smoothing averages an their standard deviation
  estAverages <- rep(x=0,times=(sampleSize/4)-2*smootheningBandwidth)
  for(k in (1:length(estAverages))){
    estAverages[k] <- mean(estimates[k:(k+2*smootheningBandwidth)])
  }
  standDevOfAvgs <- sd(estAverages)

  #test the different thresholds as per the criteria of Frahm et al.
  chosenThreshold <- 0
  for(k in (1:(length(estAverages) - plateauLength + 1)) ){
    if(sum(estAverages[(k+1):(k+plateauLength-1)] - estAverages[k]) <=
       2*standDevOfAvgs){
      chosenThreshold <- k
      break
    }
  }

  if(chosenThreshold==0){return(0)}
  else{
    return(data.frame(
      TDClogEst =
        mean(estAverages[chosenThreshold:(chosenThreshold+plateauLength-1)]),
      chosenThreshold,
      plateauLength,
      sampleSize,
      smootheningBandwidth))
  }
}

#CFG
TDCCFGEst <- function(obsMat, nBlocks=0){
  if(nBlocks == 0){
    n <- nrow(obsMat)
```

```
    U <- (1/(n+1))*rank(obsMat[,1])
    V <- (1/(n+1))*rank(obsMat[,2])
  }
  else{
    n <- nBlocks
    maxMat <- blockMaxAr(m=nBlocks, rhsVec = obsMat)
    U <- (1/(n+1))*rank(maxMat[,1])
    V <- (1/(n+1))*rank(maxMat[,2])
  }
  return( 2-2*exp( (1/n) * sum(log( ( (sqrt(log(1/U)*log(1/V))) /
                                    log((1/pmax(U,V))^2))))))
}


#FF
TDCFFEst <- function(obsMat,nBlocks=0){
  if(nBlocks == 0){
    n <- nrow(obsMat)
    U <- (1/(n+1))*rank(obsMat[,1])
    V <- (1/(n+1))*rank(obsMat[,2])
  }
  else{
    n <- nBlocks
    maxMat <- blockMaxAr(m=nBlocks, rhsVec = obsMat)
    U <- (1/(n+1))*rank(maxMat[,1])
    V <- (1/(n+1))*rank(maxMat[,2])
  }
  return(3 - (1/( 1 - (1/n)*sum(pmax(U,V)) ) ))
}

#Compilations-------------------------------------------------------------------

TDC_EVD_Ests <- function(data,nBlocks=0){
  return(data.frame(
    CFG = TDCCFGEst(data,nBlocks),
    FF = TDCFFEst(data)
  )
  )
}

TDC_d_Ests <- function(data,nBlocks=0){
  sec <- TDCsecEst(data,nBlocks)
  log <- TDClogEst(data,nBlocks)
  resultFrame <- data.frame(
    estimate = c(sec$TDCsecEst,log$TDClogEst),
    chosenThreshold = c(sec$chosenThreshold,log$chosenThreshold),
    plateauLength = c(sec$plateauLength,log$plateauLength),
    smootheningBandwidth = c(sec$smootheningBandwidth,log$smootheningBandwidth))
  rownames(resultFrame) <- c("sec","log")
```

```
  return(resultFrame)
}

TDCEsts <- function(data,nBlocks=0){
  sec <- TDCsecEst(data,nBlocks)[[1]]
  log <- TDClogEst(data,nBlocks)[[1]]
  CFG <- TDCCFGEst(data,nBlocks)
  FF <- TDCFFEst(data,nBlocks)
  return(data.frame(sec,log,CFG,FF))
}

estComp <- function(realTDC,dataAr,estSampSize=1000,sampleSize=1000,nBlocks=0){
  sortedData <- lapply(seq_len(estSampSize),function(x)dataAr[x*(1:sampleSize),])
  estimates <- do.call(rbind,
                  lapply(X=sortedData,FUN=function(x)TDCEsts(data=x,nBlocks)))

  estMeans <- colMeans(estimates)
  estBiases <- estMeans - realTDC
  estSds <- apply(X=estimates,MARGIN=2,FUN=sd)
  colNames <- c("avg","bias","sd")


  returnFrame <- data.frame(estMeans, estBiases, estSds)
  colnames(returnFrame) <- colNames

  return(returnFrame)
}

#Simulation study------------------------------------------------------------------

sampleSizes<-c(250,1000,5000)
estSampleSize <- 1000

#TDC = 0 #-------------------------------------------------------------------------

#Gaussian
cov<-0.5
sampleG <- rmvnorm(n=estSampleSize*max(sampleSizes),
                 mean=c(0,0),sigma=cbind(c(1,cov),c(cov,1)))
estimatesG <- lapply(X=sampleSizes,
                  FUN=function(x){estComp(realTDC=0,dataAr=sampleG,
                                        estSampSize = estSampleSize,
                                        sampleSize = x,
                                        nBlocks=250)})
names(estimatesG) <- paste("n =",sampleSizes)
estimatesG


#independence copula
```

```
sampleIn<- cbind(rnorm(estSampleSize*max(sampleSizes)),
               rnorm(estSampleSize*max(sampleSizes)))
estimatesIn <- lapply(X=sampleSizes,
                    FUN=function(x){estComp(realTDC=0,
                                          dataAr=sampleIn,
                                          estSampSize = estSampleSize,
                                          sampleSize = x)})
#(EV copula, so no need for block maxes)
names(estimatesIn) <- paste("n =",sampleSizes)
estimatesIn

#Frank copula
fTheta <- 1
frank <- frankCopula(param=fTheta,dim=2)
sampleF <-rCopula(max(sampleSizes)*estSampleSize,frank)
estimatesF <- lapply(X=sampleSizes,
                   FUN=function(x){estComp(realTDC=0,
                                         dataAr=sampleF,
                                         estSampSize = estSampleSize,
                                         sampleSize = x,
                                         nBlocks=250)})
names(estimatesF) <- paste("n =",sampleSizes)
estimatesF

#TDC \approx 0.5 #------------------------------------------------------------

#t-copula
roo <- 0.6045 # correlation chosen so that TDC \approx 0.5
nu <- 1.5 #degrees of freedom
Sigma <- cbind(c(1,roo),c(roo,1)) # Covariance matrix (marginals are
                              # arbitrary so the variances are set to 1)
analytic_t_TDC <- 2*pt(q= (sqrt(nu+1)*sqrt(1-roo))/sqrt(1+roo),
                     df=nu+1,lower.tail=FALSE) # \approx 0.5

samplet <- rmvt(n=max(sampleSizes)*estSampleSize,
              sigma=cbind(c(1,roo),c(roo,1)),df=nu)
estimatest <- lapply(X=sampleSizes,
                   FUN=function(x){estComp(realTDC=analytic_t_TDC,
                                         dataAr=samplet,
                                         estSampSize = estSampleSize,
                                         sampleSize = x,
                                         nBlocks=250)})
names(estimatest) <- paste("n =",sampleSizes)
estimatest

#Gumbel
gTheta <- 1/(log(x=(3/2),base=2))#chosen so that...
analytic_Gumbel_TDC <- 2-2^(1/gTheta) #... analytically TDC=0.5
gumbel <- gumbelCopula(param=gTheta,dim=2)
```

```r
sampleGum <- rCopula(max(sampleSizes)*estSampleSize,gumbel)
estimatesGum <- lapply(X=sampleSizes,
                     FUN=function(x){estComp(realTDC=analytic_Gumbel_TDC,
                                           dataAr=sampleGum,
                                           estSampSize = estSampleSize,
                                           sampleSize = x)})
#(EV copula, so no need for block maxes)
names(estimatesGum) <- paste("n =",sampleSizes)
estimatesGum

#TDC = 1 #------------------------------------------------------------------

#comonotony copula

#repeat sample twice
sampleDep <- matrix(rep(rexp(rate=1,n=max(sampleSizes)*estSampleSize),2),ncol=2)
estimatesDep <- lapply(X=sampleSizes,
                     FUN=function(x){estComp(realTDC=1,
                                           dataAr=sampleDep,
                                           estSampSize = estSampleSize,
                                           sampleSize = x)})
#(EV copula, so no need for block maxes)
names(estimatesDep) <- paste("n =",sampleSizes)
estimatesDep

# exporting into latex-----------------------------------------------------------

#plotting samples
plot(sampleIn[1:10000,],cex=0.1,pch=16,
     xlab="X",ylab="Y",main="10000 samples from In_n")
plot(sampleG[1:10000,],cex=0.1,pch=16,
     xlab="X",ylab="Y",main="10000 samples from N_n")
plot(sampleF[1:10000,],cex=0.1,pch=16,
     xlab="X",ylab="Y",main="10000 samples from F_n")
plot(samplet[1:10000,],cex=0.1,pch=16,
     xlab="X",ylab="Y",main="10000 samples from T_n")
plot(sampleGum[1:10000,],cex=0.1,pch=16,
     xlab="X",ylab="Y",main="10000 samples from Gum_n")
plot(sampleDep[1:10000,],cex=0.1,pch=16,
     xlab="X",ylab="Y",main="10000 samples from Co_n")

#exporting table
extractTable <- function(tbl){t(sapply(tbl,
                                  FUN = function(x){rbind(x$bias,x$sd)}))}
tableToExport <- extractTable(c(estimatesIn,estimatesG,estimatesF,
                                estimatest,estimatesGum,estimatesDep))
print(xtable(x=tableToExport,digits=4),
      include.rownames = F,include.colnames = F)
```

```r
#Plotting-----------------------------------------------------------------------

# More investigation (asymptotic properties of different estimators)-------------

#larger amount of sample sizes for more data points
sampleSizes<-c(2500,5000,7500,10000)
estSampleSize <- 1000

#visualise the asymptotic behaviour of the estimators when applied
# to multivariate t-data (non-extreme value data)
roo <- 0.6045 # correlation chosen so that TDC \approx 0.5
nu <- 1.5 #degrees of freedom
Sigma <- cbind(c(1,roo),c(roo,1))
# Covariance matrix (marginals are arbitrary so the variances are set to 1)
analytic_t_TDC <- 2*pt(q= (sqrt(nu+1)*sqrt(1-roo))/sqrt(1+roo),
                     df=nu+1,lower.tail=FALSE) # \approx 0.5

samplet <- rmvt(n=max(sampleSizes)*estSampleSize,
               sigma=cbind(c(1,roo),c(roo,1)),df=nu)
estimatest <- lapply(X=sampleSizes,
                   FUN=function(x){estComp(realTDC=analytic_t_TDC,
                                          dataAr=samplet,
                                          estSampSize = estSampleSize,
                                          sampleSize = x,nBlocks=250)})
names(estimatest) <- paste("n =",sampleSizes)
estimatest

#set plot colors
lineColPal <- c(viridis(4)[1:3],"#FFA500")
bgcolor <- "grey85"

#construct a data.frame for plotting with ggplot
plotFrame <- data.frame(est=rep(c("sec","log","CFG","FF"),
                             times=length(sampleSizes)*2),
                             sampleSize = rep(x=sampleSizes, each=4,times=2),
                      val = c(matrix(sapply(X=estimatest,
                                          FUN=function(x){(x[,2])}),ncol=1),
                             matrix(sapply(X=estimatest,
                                          FUN=function(x){(x[,3])}),ncol=1)),
                      meas = c(rep(x="bias",times=4*length(sampleSizes)),
                              rep(x="sd",times=4*length(sampleSizes)))
)

#plot
basePlot <- ggplot() + scale_color_manual(values = lineColPal) +
  geom_line(data=plotFrame,
           aes(x=sampleSize,y=val,color=est,linetype=meas),size=1.2)+
  geom_point(data=plotFrame,
            aes(x=sampleSize,y=val,color=est,shape=meas),size=3)
```

```
basePlot <- basePlot + scale_x_continuous(breaks=sampleSizes) +
  xlab("n") + ylab("") + ggtitle("Estimator performance on T_n")
basePlot +
  theme(panel.background = element_rect(fill=bgcolor,colour=bgcolor,
                                        size=0.5,linetype="solid")
        ,legend.box="horizontal",legend.position ="bottom",
        legend.background = element_rect(fill="white",colour="black"),
        plot.title = element_text(hjust = 0.5),
        legend.title=element_blank())

#visualise asymptotic beahviour of estimators when applied to
# data simulated from the Gumbel copula (extreme value data)
gTheta <- 1/(log(x=(3/2),base=2))#chosen so that...
analytic_Gumbel_TDC <- 2-2^(1/gTheta) #... analytically TDC=0.5
gumbel <- gumbelCopula(param=gTheta,dim=2)
sampleGum <- rCopula(max(sampleSizes)*estSampleSize,gumbel)
estimatesGum <- lapply(X=sampleSizes,
                     FUN=function(x){estComp(realTDC=analytic_Gumbel_TDC,
                                             dataAr=sampleGum,
                                             estSampSize = estSampleSize,
                                             sampleSize = x)})
names(estimatesGum) <- paste("n =",sampleSizes)
estimatesGum

#data.frame for plotting with ggplot
plotFrame <- data.frame(est=rep(c("sec","log","CFG","FF"),
                             times=length(sampleSizes)*2),
                             sampleSize = rep(x=sampleSizes, each=4,times=2),
                      val = c(matrix(sapply(X=estimatesGum,
                                          FUN=function(x){(x[,2])}),ncol=1),
                             matrix(sapply(X=estimatesGum,
                                          FUN=function(x){(x[,3])}),ncol=1)),
                      meas = c(rep(x="bias",times=4*length(sampleSizes)),
                              rep(x="sd",times=4*length(sampleSizes)))
)

#plotting
basePlot <- ggplot() + scale_color_manual(values = lineColPal) +
  geom_line(data=plotFrame,
           aes(x=sampleSize,y=val,color=est,linetype=meas),size=1.2)+
  geom_point(data=plotFrame,
            aes(x=sampleSize,y=val,color=est,shape=meas),size=3)
basePlot <- basePlot + scale_x_continuous(breaks=sampleSizes) +
  xlab("n") + ylab("") + ggtitle("Estimator performance on Gum_n")
basePlot +
  theme(panel.background = element_rect(fill=bgcolor,colour=bgcolor,
                                        size=0.5,linetype="solid")
        ,legend.box="horizontal",legend.position ="bottom",
        legend.background = element_rect(fill="white",colour="black"),
```

```
        plot.title = element_text(hjust = 0.5),
        legend.title=element_blank())

#Testing for performance in increasingly tail dependent samples------------------

#generate samples from Gumbel copula with varying parameter
sampleSize<-1000
estSampleSize<-1000
gThetas <- 1/(log(x=(c(14,12,10,81/10)/8),base=2))
analytic_Gumbel_TDCs <- 2-2^(1/gThetas)
gumbelCops <- lapply(X=gThetas,FUN=function(x){gumbelCopula(param=x,dim=2)})
samplesGum <- lapply(X=gumbelCops,
                   FUN=function(x){rCopula(sampleSize*estSampleSize,x)})
names(samplesGum) <- analytic_Gumbel_TDCs
estimatesGum <- lapply(X=seq_along(samplesGum),
                     FUN=function(x,y,i){estComp(realTDC=y[i],dataAr=x[[i]],
                                                 estSampSize = estSampleSize,
                                                 sampleSize = sampleSize)}
                     , x=samplesGum, y = analytic_Gumbel_TDCs)
names(estimatesGum) <- paste("TDC =",analytic_Gumbel_TDCs)
estimatesGum

#export as table
tableToExport <- sapply(X=estimatesGum,FUN=function(x){rbind(x$bias,x$sd)})
print(xtable(x=tableToExport,digits=4),include.rownames = F,include.colnames = T)
#plot as well
dataFrameToPlot <- data.frame( TDC = rep(x=analytic_Gumbel_TDCs,
                                     each=length(analytic_Gumbel_TDCs)*2),
                            est = rep(x=rep(c("sec","log","CFG","FF"),each=2),
                                      times=length(analytic_Gumbel_TDCs)),
                            val = matrix(sapply(X=estimatesGum,
                                               FUN=function(x){rbind(x$bias,
                                                                     x$sd)}),
                                         ncol=1),
                            meas = rep(x=c("bias","sd"),
                                     times=2*4*length(analytic_Gumbel_TDCs)))

basePlot <- ggplot()+ scale_color_manual(values = lineColPal) +
  geom_line(data=dataFrameToPlot,
           aes(x=TDC,y=val,color=est,linetype=meas),size=1.2)+
  geom_point(data=dataFrameToPlot,aes(x=TDC,y=val,color=est,shape=meas),size=3)

basePlot <- basePlot + scale_x_continuous(breaks=analytic_Gumbel_TDCs) +
  xlab("TDC") + ylab("") +
  ggtitle("Estimator performance on Gumbel distributions with varying TDC")

basePlot +
  theme(panel.background = element_rect(fill=bgcolor,colour=bgcolor,
                                    size=0.5,linetype="solid")
```

```
        ,legend.box="horizontal",legend.position ="bottom",
        legend.background = element_rect(fill="white",colour="black"),
        plot.title = element_text(hjust = 0.5),
        legend.title=element_blank())

#try out the t-distribution as well
require("mvtnorm")
sampleSize<-1000
estSampleSize<-1000
roos <- c(0.1,0.6045,0.9,0.9999)
nu <- 1.5 #degrees of freedom
analytic_t_TDCs <- 2*sapply(X=roos,
                    FUN=function(x){pt(q= (sqrt(nu+1)*sqrt(1-x))/sqrt(1+x)
                                     ,df=nu+1,lower.tail=FALSE)})

samplest <- lapply(X=roos,
                 FUN=function(x){rmvt(n=sampleSize*estSampleSize,
                                  sigma=cbind(c(1,x),c(x,1)),df=nu)})
estimatest <- lapply(X=seq_along(samplest),
                   FUN=function(i){estComp(realTDC=analytic_t_TDCs[i],
                                       dataAr=samplest[[i]],
                                       estSampSize = estSampleSize,
                                       sampleSize = sampleSize,
                                       nBlocks=250)})
names(estimatest) <- paste("TDC =",analytic_t_TDCs)
estimatest

#Frequency plots for estimates----------------------------------------------------
freqPlotEst <- function(realTDC,dataAr,estSampSize=1000,
                      sampleSize=1000,nBlocks=0,type,title){
  #chop the data up into individual samples
  sortedData <- lapply(seq_len(estSampSize),
                     function(x)dataAr[x*(1:sampleSize),])
  #use the "type"-variable to decide upon an estimation method
  estFun <- function(data,nBlocks){switch(type,
                                     "sec" = TDCsecEst(data,nBlocks)[[1]],
                                     "log" = TDClogEst(data,nBlocks)[[1]],
                                     "CFG" = TDCCFGEst(data,nBlocks),
                                     "FF" = TDCFFEst(data,nBlocks))}
  #calculate estimates
  estimates <- do.call(rbind,
                     lapply(X=sortedData,
                           FUN=function(x){estFun(data=x,nBlocks=nBlocks)}))
  meanEst <- mean(estimates)
  sdEst <- sd(estimates)

  #set color palette
  #colPal <- viridis(max(ceiling((abs(estimates - meanEst)/sdEst)))+1)
  colPal <- rev(brewer.pal(n=max(ceiling((abs(estimates - meanEst)/sdEst))+1,
```

```
                         name="Blues"))
  bgcolour <- "blue"

  #frequency plotting
  plotData <- data.frame(val = estimates,
                       sdDist = ceiling(abs((estimates - meanEst)/sdEst)))
  basePlot <- ggplot(data=plotData,aes(x=val,fill=as.factor(sdDist))) +
    scale_fill_manual(values = colPal) +
    geom_histogram(bins = 100,color="black")

  basePlot +
    geom_vline(xintercept = realTDC,linetype="solid",color="red",size=1.2)+
    geom_vline(xintercept = meanEst,linetype="solid",color="orange",size=1.2)+
    geom_vline(xintercept = c(meanEst-sdEst,meanEst+sdEst),
               linetype="dashed",color=colPal[2],size=1)+
    geom_vline(xintercept = c(meanEst-2*sdEst,meanEst+2*sdEst),
               linetype="dashed",color=colPal[3],size=1)+
    xlab("estimator value")+ylab("frequency")+ggtitle(title)+
    theme(panel.background = element_rect(fill=bgcolor,colour=bgcolor,
                                          size=0.5,linetype="solid")
          ,legend.box="horizontal",legend.position ="bottom",
          legend.background = element_rect(fill="white",colour="black"),
          plot.title = element_text(hjust = 0.5))+
    scale_y_continuous(expand = c(0,0),limits=c(0,35)) +
    guides(fill =
             guide_legend("Number of standard deviations\naway from the mean"))

}

#Example of bad performance
sampleSize<-5000
estSampleSize<-1000
cov<-0.5
sampleG <- rmvnorm(n=estSampleSize*sampleSize,mean=c(0,0),
                   sigma=cbind(c(1,cov),c(cov,1)))
freqPlotEst(realTDC=0,dataAr=sampleG,estSampSize=estSampleSize,
            sampleSize=sampleSize,nBlocks=250,
            type="log",
            title=
      "1000 realisations of the log estimator\n when applied to sample N_5000")

#Example of good performance
sampleSize<-5000
estSampleSize<-1000
gTheta <- 1/(log(x=(3/2),base=2))
analytic_Gumbel_TDC <- 2-2^(1/gTheta)
gumbelCop <- gumbelCopula(param=gTheta,dim=2)
sampleGum <- rCopula(sampleSize*estSampleSize,gumbelCop)
freqPlotEst(realTDC=analytic_Gumbel_TDC,dataAr=sampleGum,
```

```
            estSampSize=estSampleSize,sampleSize=sampleSize,nBlocks=0,
            type="FF",
            title=
      "1000 realisations of the FF estimator when applied to sample Gum_5000")

#Copula plots-----------------------------------------------------------------

#plot copulas in order of appearance in the thesis:
# independence copula (Gumbel with theta =1 produces the ind.cop.),
# comonotonic copula (Gumbel copula -> comonotonic copula as theta -> infty),
# Frank copula (theta = -5),
# Gumbel copula (theta = 1.5),
# Gaussian copula (cov = 0.1).
# The choises of parametrisation have been made to exaggerate the shape of
# the graphs.

#choose color palette
regColPal <- colorRampPalette(c("blue","#00FFFF"))(100)
lineColor <-"white"

#surface plots
lapply(X=c(gumbelCopula(1),gumbelCopula(250),frankCopula(-5),
          gumbelCopula(1.5),normalCopula(0.1)),
      FUN=function(y){
        wireframe2(x=y,FUN=pCopula,
                  drape=TRUE,colorkey=FALSE,lwd=1,col=lineColor,
                  alpha.regions=1,col.regions=regColPal,
                  n.grid=15,draw.4.pCoplines = FALSE,
                  xlab="u",ylab="v",zlab="C(u,v)",
                  zoom=1.05)
      }
)

#contour plots
lapply(X=c(gumbelCopula(1),gumbelCopula(250),frankCopula(-5),
          gumbelCopula(1.5),normalCopula(0.1)),
      FUN=function(y){
        contourplot2(x=y,FUN=pCopula,colorkey=FALSE,lwd=2,
                    col=lineColor,col.regions=regColPal,
                    xlab="u",ylab="v",zlab="C(u,v)")
      }
)
```

References

[1] Søren Asmussen and Mogens Steffensen. *Risk and Insurance: A Graduate Text.* Vol. 96. Springer Nature, 2020.

[2] Philippe Capéraà, A-L Fougères, and Christian Genest. "A nonparametric estimation procedure for bivariate extreme value copulas". In: *Biometrika* 84.3 (1997), pp. 567–577.

[3] Fabrizio Durante, Juan Fernandez-Sanchez, and Carlo Sempi. "A topological proof of Sklar's theorem". In: *Applied Mathematics Letters* 26.9 (2013), pp. 945–948.

[4] Paul Embrechts, Filip Lindskog, and Alexander McNeil. "Modelling dependence with copulas". In: *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich* 14 (2001).

[5] Paul Embrechts, Alexander McNeil, and Daniel Straumann. "Correlation and dependence in risk management: properties and pitfalls". In: *Risk management: value at risk and beyond* 1 (2002), pp. 176–223.

[6] Ana Ferreira and Laurens De Haan. "On the block maxima method in extreme value theory: PWM estimators". In: *The Annals of statistics* 43.1 (2015), pp. 276–298.

[7] Helena Ferreira and Marta Ferreira. "On extremal dependence of block vectors". In: *Kybernetika* 48.5 (2012), pp. 988–1006.

[8] Marta Susana Ferreira. "Nonparametric estimation of the tail-dependence coefficient". In: (2013).

[9] Gabriel Frahm, Markus Junker, and Rafael Schmidt. "Estimating the tail-dependence coefficient: properties and pitfalls". In: *Insurance: mathematics and Economics* 37.1 (2005), pp. 80–100.

[10] Christian Genest and Johan Segers. "Rank-based inference for bivariate extreme-value copulas". In: *The Annals of Statistics* 37.5B (2009), pp. 2990–3022.

[11] Robert D Gordon. "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument". In: *The Annals of Mathematical Statistics* 12.3 (1941), pp. 364–366.

[12] Gordon Gudendorf and Johan Segers. "Extreme-value copulas". In: *Copula theory and its applications.* Springer, 2010, pp. 127–145.

[13] Petri Koistinen. *Todennäköisyyslaskenta.* 2013.

[14] Jaakko Lehtomaa. *Risk Theory.* 2019.

[15] Roger B Nelsen. *An introduction to copulas.* Springer Science & Business Media, 2007.

[16] Sidney I Resnick. *Extreme values, regular variation and point processes.* Springer, 2013.

[17] Rafael Schmidt and Ulrich Stadtmüller. "Non-parametric estimation of tail dependence". In: *Scandinavian Journal of Statistics* 33.2 (2006), pp. 307–335.

[18] Johan Segers. "Non-parametric inference for bivariate extreme-value copulas". In: (2004).

[19] Aad W Van der Vaart. *Asymptotic statistics.* Vol. 3. Cambridge university press, 2000.

[20] Aad W Van Der Vaart et al. *Weak convergence and empirical processes: with applications to statistics.* Springer Science & Business Media, 1996.