Universidad Autónoma de Madrid

Facultad de Ciencias
Escuela Politécnica Superior

Modelling Multivariate Dependencies with Semiparametric Archimedean Copulas

Master's thesis

In partial fulfilment of the requirements
for the joint master's degree in:

Matemáticas y Aplicaciones
Investigación e Innovación en TIC

Author:

Javier Fernández Serrano

Advisors:

Dr. Alberto Suárez González
Dr. Santiago Carrillo Menéndez

July, 2016

A mi madre y a mi novia

Agradecimientos

En primer lugar, quiero agradecer a mis tutores, los doctores Alberto Suárez González y Santiago Carrillo Menéndez, el haberme permitido elegir el tema de este trabajo, así como su dedicación y paciencia durante todo este tiempo.

En segundo lugar, quiero dar las gracias a la Universidad Autónoma de Madrid por la formación que me ha dado, brindándome la oportunidad de simultanear las carreras de matemáticas y de informática, tanto a nivel de licenciatura e ingeniería como de máster.

Por último, quiero agradecer a la profesora Dra. Silvia T. Acuña Castillo el haberme iniciado en la investigación y todo el entusiasmo y esmero que puso en mi formación.

Javier Fernández Serrano
Julio, 2016

Abstract

While there exists a vast repertoire of probability distributions and estimation methods in the literature to model univariate random variables, multivariate models remain comparatively less developed. Despite the breakthrough of copulas in the late 1950s and their rapid development in the early 2000s, the quest for flexible models in dimensions higher than two goes on to this day. This work contributes to this goal by extending a successful semiparametric Archimedean bivariate copula estimation method to the 3-variate case. Our approach is based on tensor product splines, conditional copulas and vine constructions. A novel regularization procedure to reduce overfitting is also proposed. Experiments with simulated data show that the proposed model can represent complex dependencies expressed in terms of Kendall's tau, a dependence measure that is made to vary smoothly with a conditioning variable.

Keywords: copula, Archimedean copula, conditional copula, vine, multivariate statistics, semiparametric models, spline models, regularization, Kendall's tau, tail dependence

Notational preliminaries

Throughout this work, the acronyms 'r.v.', 'CDF' and 'pdf' shall serve as shorthand for 'random variable', 'cumulative distribution function' and 'probability density function', respectively. Most of the times, we will be using CDFs. Notwithstanding, when dealing with r.v.'s, we may assume a generic underlying probability space $(\Omega, \Sigma, \mathbb{P})$.

Algebra

$(\mathbf{x}; \mathbf{y})$	The $(n+m)$-dimensional vector resulting from stacking vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$.
$\operatorname{Ker} A$	The kernel subspace of the matrix or linear map A. .
$\mathcal{M}_{n\times m}(\mathbb{R})$	Real matrices with n rows and m columns .
$\mathcal{P}^d(\mathbb{R})$	Polynomials with real coefficients of degree d .
$\operatorname{vec} A$	The vectorization of matrix A, i.e., the (column) vector resulting from stacking the columns of A .
$A \otimes B$	The Kronecker product of matrices A and B .

Calculus

$\mathcal{C}(\Omega)$	The set of all continuous functions $f : \Omega \to \mathbb{R}$.
$\mathcal{C}^k(\Omega)$	The set of all functions $f : \Omega \to \mathbb{R}$ with continuous k-th order partial derivatives. If $k = \infty$, partial derivatives of any order are assumed to exist .
$\mathcal{C}^k(a, b)$	The set of all functions $f : (a, b) \to \mathbb{R}$ with continuous k-th order derivative. If $k = \infty$, derivatives of any order are assumed to exist .
$x \to x_0^+$	x approaches x_0 from the right $(x_0 < x)$.
$x \to x_0^-$	x approaches x_0 from the left $(x < x_0)$.

Functions

$\operatorname{Im}(f)$	The image of function f. .
$\mathbb{1}_A$	The function mapping x to 1 if $x \in A$ or to 0, otherwise .
$x \mapsto f(x)$	The anonymous function mapping x to $f(x)$.

Probability and statistics

$[X\|Y=y]$	The r.v. that follows the conditional distribution of r.v. X given that the r.v. Y is equal to y .
$\mathbb{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$	The n-dimensional r.v. $\mathbb{X}$ follows an n-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ .
$\mathbb{X} \sim t_n(\nu, \boldsymbol{\mu}, \Sigma)$	The n-dimensional r.v. $\mathbb{X}$ follows an n-variate Student's t distribution with ν degrees of freedom, location parameter μ and scale parameter σ^2 .
$\mathcal{Q}_X(p)$	The p-th quantile, $p \in [0,1]$, of the r.v. X .
$F^{\leftarrow}(p)$	The p-th quantile, $p \in [0,1]$, of the CDF F .
$X = Y$ a.s.	The r.v. X is almost surely equal to the r.v. Y, that is $\mathbb{P}(X \neq Y) = 0$.
$X \overset{d}{=} Y$	The r.v. X follows the same distribution as the r.v. Y .
$X \sim \mathcal{N}(\mu, \sigma^2)$	The r.v. X follows a univariate normal distribution with mean μ and variance σ^2 .
$X \sim \mathcal{U}[0,1]$	The r.v. X follows a uniform distribution in $[0,1]$.
$X \sim F$	The r.v. X is distributed according to the CDF F .
$X \sim t(\nu, \mu, \sigma^2)$	The r.v. X follows a Student's t distribution with mean μ and variance σ^2 .

Sets and numbers

$\#A$	The number of elements in the finite set A .
$\bar{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$.
$\mathbb{N}_n$	$\{1, 2, \ldots, n\}$.
$\mathbb{Z}_+$	$\{n \in \mathbb{Z} : n \geq 0\}$.
$\mathbf{0}_n$	n-dimensional null vector .

Contents

List of figures

Chapter 1

Introduction

> Implícita en todas estas definiciones está la suposición de que el número de humanos es suficientemente grande para un tratamiento estadístico válido. El tamaño necesario de tal número puede ser determinado por el primer teorema de Seldon, que... Otra suposición necesaria es que el conjunto humano debe desconocer el análisis psicohistórico a fin de que su reacción sea verdaderamente casual...
>
> *Fundación*
> Isaac Asimov

There exists a vast repertoire of univariate probability distributions in the literature. Depending on the desired application, one can often assume a model that belongs to a specific parametric family which is known to be adequate in similar scenarios. By contrast, the theory of multivariate distributions is less developed. One of the best-known multivariate distribution families is that of elliptical distributions (Cambanis et al., 1981). However, they are often too rigid in practice. For instance, the univariate margins and conditional margins of elliptical distributions are always elliptical.

The introduction in 1959 of the concept of *copula* (Sklar, 1973) by Sklar is a breakthrough in the modelling of dependencies. Copulas are multivariate distributions whose univariate margins are uniformly distributed in $[0, 1]$. Copulas allow to separate the modelling of dependencies from the modelling of marginals, according to Sklar's theorem, and appear as a useful theoretical tool to study dependence measures such as Kendall's tau, Spearman's rho or the upper and lower tail dependence indices.

Despite the interesting mathematical properties of copulas, it was not until the 2000s that copulas drew the attention of both researchers and practitioners, especially of those interested in risk management and finance (Bouyé et al., 2000), (Embrechts et al., 2003), (McNeil et al., 2005). In the early days of copula theory, researchers focused on expanding the set of known copulas, mainly introducing parametric bivariate copulas (dimension $n = 2$); nowadays, the focus is on obtaining flexible multivariate copula models (dimension $n > 2$) capable of representing complex dependencies from data.

The goal of this work is to make headway in both the flexibility and the dimensionality aspects of copula modelling. In this work, we will focus on the

$n = 2$ and $n = 3$ cases. According to Lopez-Paz et al.:

> "Although there exist many parametric models for two-dimensional copulas, for more than two dimensions the number and expressiveness of families of parametric copulas is more limited."
>
> (Lopez-Paz et al., 2013)

To achieve this goal, we take (Hernández-Lobato and Suárez, 2011) as the starting point for our proposal. (Hernández-Lobato and Suárez, 2011) introduces a family of semiparametric Archimedean copulas, which specifically tackles tail dependence estimation. We propose to extend (Hernández-Lobato and Suárez, 2011) to conditional copulas in a similar way to the approach in (Lambert, 2014). Finally, we propose to build 3-variate copula models from bivariate copulas and a family of conditional copulas by means of the vine constructions (Bedford and Cooke, 2002).

Our work is organized as follows. Chapter 2 introduces the essentials of copula theory. Special attention is given to various dependence measures (Section 2.3), Archimedean copulas (Section 2.4.2) and vines (Section 2.6). Chapter 3 reviews the state of the art of copula modelling, with some emphasis on the work by Lambert (Section 3.2.3). Chapter 4 describes the bivariate semiparametric model introduced by (Hernández-Lobato and Suárez, 2011) (Section 4.1), which is the starting point for our proposal. This method is extended to 3-copulas in Section 4.2. Chapter 5 presents the results of simulations that illustrate the results in Chapter 4. Finally, Chapter 6 summarizes the conclusions of this work and outlines future lines of research.

Appendices A, B, C are self-contained summaries of relevant topics for statistical modelling: spline functions, regular variation and quantile functions. Appendix D, which addresses implementation details, provides further details that are relevant to the topics discussed in Chapter 4.

Chapter 2

Fundamentals of copula theory

> – ¿Matemáticas? – exclamó Quemot, pronunciando con voz de falsete la última sílaba.
> – Verá, no las matemáticas superiores que se emplean en robótica, que sería incapaz de entender, sino matemáticas aplicadas a la sociología. Por ejemplo, me es muy familiar la Relación de Teramin.
> – ¿La qué?
> – Acaso ustedes la conozcan por un nombre distinto. El diferencial de vejaciones sufridas con provilegios concedidos: D a sub J elevado a la enésima...
> – ¿De qué está usted hablando?
>
> *El sol desnudo*
> Isaac Asimov

In this chapter we will present a brief introduction to copula theory. We will summarize some basic definitions and properties and introduce families of copulas that will be used throughout this work.

This review of copulas and their properties is included for the sake of completeness and as a quick reference. Notwithstanding, we shall not focus on demonstrations in this chapter and refer the reader to the bibliography for further details. Probably two of the most complete works on this topic at an introductory level are (Embrechts et al., 2003) and (Nelsen, 2006). The latter stands out for its rigorous approach. In the context of finance and risk management, (Bouyé et al., 2000) and (McNeil et al., 2005) provide good introductions to copulas and their applications, too.

2.1 Basic definitions and properties

We now introduce the concept of copula in its most general form, even though we shall only deal with the $n = 2$ and $n = 3$ cases in this work.

Definition 2.1.1 (Copula). A function $C : [0,1]^n \to \mathbb{R}$ is called an n-variate copula or an n-copula, for $n > 1$, if the following hypotheses are met:

1. For all $u_1, u_2, \ldots, u_n \in [0,1]$, if $u_i = 0$ for some $i = 1, 2, \ldots, n$, then $C(u_1, u_2, \ldots, u_n) = 0$.

2. For all $u \in [0,1]$,

$$C(u, 1, 1, \ldots, 1) = C(1, u, 1, 1, \ldots, 1) = \ldots = C(1, 1, \ldots, 1, u) = u \ .$$

3. For every non-empty rectangle $R = (a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_n, b_n] \subset (0,1]^n$, we have $\mu_C(R) \geq 0$, where

$$\mu_C(R) = \sum_{\mathbf{u} \in V(R)} \operatorname{sign}(\mathbf{u}) \cdot C(\mathbf{u}) \ , \tag{2.1}$$

for the vertices set $V(R) = \{a_1, b_1\} \times \{a_2, b_2\} \times \ldots \times \{a_n, b_n\}$ and with the sign of a vertex $\mathbf{u} \in V(R)$ defined as $\operatorname{sign}(\mathbf{u}) = (-1)^{N(\mathbf{u})}$, for $N(u_1, \ldots, u_n) = \#\{i \in \mathbb{N}_n : u_i = a_i\}$.

From the previous definition one can easily see that a copula is in fact the cumulative distribution function (CDF) of an n-variate random vector (r.v.) with uniform margins in $[0,1]$ restricted to $[0,1]^n$. That is, if $U_1, U_2, \ldots, U_n \sim \mathcal{U}[0,1]$, then

$$C(u_1, u_2, \ldots, u_n) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_n \leq u_n) \tag{2.2}$$

is an n-copula. The third axiom in Definition 2.1.1, which states that an n-copula is 'n-increasing', guarantees that C produces a (probability) measure by means of (2.1), while the second one implies the margins are uniform in $[0,1]$. Some authors (Nelsen, 2003) restrict the range of a copula to $[0,1]$, but this is actually implied by Definition 2.1.1.

Equation (2.1) seems a bit intimidating in the n-variate version. For the $n = 2$ case, we can simply write

$$\mu_C\left((a,b] \times (c,d]\right) = C(b,d) - C(a,d) - C(b,c) + C(a,c) \geq 0 \ .$$

Definition 2.1.1 entails a considerable degree of regularity. In particular, every copula is Lipschitz-continuous with Lipschitz constant $L = 1$ and, hence, uniformly continuous.

Proposition 2.1.1. *Let C be an n-copula and let $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in [0,1]^n$ and $\mathbf{u}' = (u_1', u_2', \ldots, u_n') \in [0,1]^n$. We have*

$$|C(\mathbf{u}) - C(\mathbf{u}')| \leq \|\mathbf{u} - \mathbf{u}'\|_1 = \sum_{i=1}^{n} |u_i - u_i'| \ .$$

Remark 2.1.1. For Rademacher's theorem, since an n-copula C is Lipschitz continuous (Proposition 2.1.1), C is differentiable almost everywhere in $(0,1)^n$.

(Schmitz, 2003) provides an extensive and rigorous treatment of the partial derivatives of copulas.

Proposition 2.1.2. *Let C be an n-copula and let $i \in \mathbb{N}_n$. For all $u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n \in (0,1)$, the function*

$$u_i \mapsto \frac{\partial C}{\partial u_i}(u_1, u_2, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_n) \ , \tag{2.3}$$

is defined and monotonically increasing for almost every u_i *in* $(0,1)$*. Moreover, for almost every* $(u_1, u_2, \ldots, u_n) \in (0,1)^n$*, we have*

$$0 \leq \frac{\partial C}{\partial u_i}(u_1, u_2, \ldots, u_n) \leq 1 \ ,$$

and, in the context of (2.2), the following interpretation holds:

$$\frac{\partial C}{\partial u_i}(u_1, u_2, \ldots, u_n) =$$
$$\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_{i-1} \leq u_{i-1}, U_{i+1} \leq u_{i+1}, \ldots, U_n \leq u_n | U_i = u_i) \ .$$

Next we will introduce three important copulas that correspond to three basic types of dependence. Strictly speaking, they are copula *templates* that can be instantiated for every dimension n. For every n, two of these copulas act as *bounds* for the whole set of n-copulas, according to the following result.

Proposition 2.1.3 (**Fréchet-Hoeffding bounds**). *Let* C *be an* n*-copula. For all* $u_1, u_2, \ldots, u_n \in [0,1]$*, we have*

$$\max\left\{1 - n + \sum_{i=1}^{n} u_i, 0\right\} \leq C(u_1, u_2, \ldots, u_n) \leq \min\{u_1, u_2, \ldots, u_n\} \ . \quad (2.4)$$

The upper bound in (2.4) is actually the CDF of the n-variate r.v. $(U, U, \ldots, U)$ for a r.v. $U \sim \mathcal{U}[0,1]$.

Definition 2.1.2 (**Maximal copula**). We define the n-variate maximal copula $C^+ : [0,1]^n \to \mathbb{R}$ as

$$C^+(u_1, \ldots, u_n) = \min\{u_1, u_2, \ldots, u_n\} \ .$$

By contrast, the lower bound in (2.4) is *not* a copula in general, but a related concept known as quasi-copula[1]. It is a copula if and only if $n = 2$, in which case the copula corresponds to the bivariate r.v. $(U, 1-U)$, for a r.v. $U \sim \mathcal{U}[0,1]$.

Definition 2.1.3 (**Minimal copula**). We define the minimal copula $C^- : [0,1]^2 \to \mathbb{R}$ as

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

Last but not least, we present the independence copula. As its name suggests, it is the CDF of the multivariate r.v. whose univariate margins are uniformly distributed in $[0,1]$ and mutually independent.

Definition 2.1.4 (**Independence copula**). The n-variate independence copula (also known as product copula) $C^\perp : [0,1]^n \to \mathbb{R}$ is defined as

$$C^\perp(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i \ .$$

[1] We shall not expand on quasi-copulas. See (Nelsen, 2003) for further details.

Kendall distribution

Given a copula, we can always define the following concept, which will be useful later on.

Definition 2.1.5 (**Kendall distribution**). Let C be a bivariate copula. Let (U, V) be uniform r.v.'s in $[0, 1]$ jointly distributed according to C. We define the Kendall distribution of the copula C as

$$K_C(x) = \mathbb{P}(C(U, V) \leq x) \ . \tag{2.5}$$

2.2 Sklar's theorem

The importance of copulas in statistics is due to Sklar's theorem. This famous result states that, under certain hypotheses, a multivariate distribution can be broken down into two different kinds of structures: a unique copula on the one hand and the univariate margins on the other. This poses strong implications for statistical modelling, as it suggests that the estimation of dependence in a multivariate distribution comes down to the estimation of a copula.

Theorem 2.2.1 (**Sklar's theorem, 1959**). *Let F be an n-variate CDF, $n > 1$, with univariate margins $F_1, F_2, \ldots, F_n$. There exists an n-copula C such that*

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \ .$$

Moreover, if the F_i's are continuous, then C is unique.

Using Proposition C.0.12(4), if the F_i's are continuous, the unique copula C in Theorem 2.2.1 is given by

$$C(u_1, u_2, \ldots, u_n) = F(F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2), \ldots, F_n^{\leftarrow}(u_n)) \ , \tag{2.6}$$

where the $F_i^{\leftarrow}$'s are the quantile[2] functions corresponding to the CDFs F_i's. Interestingly, the underlying copula, in the sense of Sklar's theorem, of a multivariate r.v. remains invariant under strictly increasing transformations, as the following result states.

Proposition 2.2.1. *Let $X_1, X_2, \ldots, X_n$ be unidimensional continuous r.v.'s. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be strictly increasing functions on $\mathrm{Im}(X_1), \mathrm{Im}(X_2), \ldots, \mathrm{Im}(X_n)$, respectively. We have that the n-variate r.v.'s $(X_1, X_2, \ldots, X_n)$ and $(\alpha_1(X_1), \alpha_2(X_2), \ldots, \alpha_n(X_n))$ share the same associated copula, in the sense of Sklar's Theorem* 2.2.1.

The previous result allows to obtain a stochastic representation of the n-variate r.v. (with uniform margins in $[0, 1]$) distributed according to the copula of n unidimensional r.v.'s.

Corollary 2.2.1. *Let $X_1, X_2, \ldots, X_n$ be unidimensional continuous r.v.'s. with corresponding CDFs $F_1, F_2, \ldots, F_n$. We have that the n-variate r.v.'s $(X_1, X_2, \ldots, X_n)$ and $(F_1(X_1), F_2(X_2), \ldots, F_n(X_n))$ share the same associated copula C, which is also the multivariate CDF of the latter r.v., restricted to $[0, 1]^n$.*

[2] See Appendix C for further details.

Proof. It is a consequence of (2.2.1), noting that we can always find a version of X_i such that F_i is strictly increasing in $\mathrm{Im}(X_i)$.[3] □

The converse of Sklar's theorem also holds: given a copula and some univariate CDFs, we can construct a multivariate CDF. While Sklar's Theorem 2.2.1 allows to extract the dependence structure from a given known multivariate CDF, its converse, Proposition 2.2.2, allows to create new CDFs with a given dependence structure. This represents a big asset, since the number of known multivariate CDFs is relatively small, whereas constructing new copulas is a simpler task.

Proposition 2.2.2 (**Multivariate CDF construction via copulas**). *Let C be an n-copula and let $F_1, F_2, \ldots, F_n$ be n univariate CDFs. Letting $F : \mathbb{R}^n \to [0,1]$ be defined as*

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \ ,$$

we have that F is an n-variate CDF whose i-th univariate margin coincides with F_i, for all $i = 1, 2, \ldots, n$.

There exists a version of Sklar's Theorem 2.2.1 involving *survival* functions (McNeil et al., 2005). If we define F as the n-variate CDF

$$F(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \ ,$$

for some r.v.'s $X_1, X_2, \ldots, X_n$, then the survival function $\bar{F}$ corresponding to F is defined as

$$\bar{F}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 > x_1, X_2 > x_2, \ldots, X_n > x_n) \ .$$

Similarly, if $F_i(x) = \mathbb{P}(X_i \leq x)$ is the i-th univariate margin of F, its corresponding survival function is $\bar{F}_i(x) = \mathbb{P}(X_i > x)$.

Theorem 2.2.2 (**Sklar's theorem, survival version**). *Let $\bar{F}$ be an n-variate survival function, $n > 1$, with univariate margin survival functions $\bar{F}_1, \bar{F}_2, \ldots, \bar{F}_n$. There exists an n-copula $\hat{C}$ such that*

$$\bar{F}(x_1, x_2, \ldots, x_n) = \hat{C}\left(\bar{F}_1(x_1), \bar{F}_2(x_2), \ldots, \bar{F}_n(x_n)\right) \ .$$

Moreover, if the $\bar{F}_i$'s are continuous, then $\hat{C}$ is unique.

$\hat{C}$ is known as the survival *copula* of C, which is not to be confused with the survival *function* $\bar{C}$ of C.[4] Letting $U_i = F_i(X_i)$ and $\bar{U}_i = 1 - U_i$, for all $i = 1, 2, \ldots, n$, one can show that $\hat{C}$ is actually the CDF of $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_n)$. Therefore, the relationship between $\hat{C}$ and $\bar{C}$ is:

$$\begin{aligned} \hat{C}(1-u_1, 1-u_2, \ldots, 1-u_n) &= \mathbb{P}(\bar{U}_1 \leq 1-u_1, \bar{U}_2 \leq 1-u_2, \ldots, \bar{U}_n \leq 1-u_n) \\ &= \mathbb{P}(U_1 \geq u_1, U_2 \geq u_2, \ldots, U_n \geq u_n) \\ &= \mathbb{P}(U_1 > u_1, U_2 > u_2, \ldots, U_n > u_n) \\ &= \bar{C}(u_1, u_2, \ldots, u_n) \end{aligned} \ ,$$

for all $u_1, u_2, \ldots, u_n \in [0,1]$.

The property $C = \hat{C}$ deserves its own definition.

[3] If F_i is flat on $I = \cup_{j=1}^{n} I_j$, then $\mathbb{P}(X_i \in I) = 0$. Therefore, we can find $\tilde{X}_i$ that is equal to X_i almost surely such that F_i is strictly increasing on $\mathrm{Im}(\tilde{X}_i)$.

[4] C is, in particular, a multivariate CDF and, hence, has a survival function.

Definition 2.2.1 (**Radially symmetric copula).** An n-copula C is said to be radially symmetric if it is equal to its survival copula $\hat{C}$, that is, if $C(u_1, u_2, \ldots, u_n) = \hat{C}(u_1, u_2, \ldots, u_n)$ for all $u_1, u_2, \ldots, u_n \in [0, 1]$.

In the bivariate case, $n = 2$, there is also a simple relationship between the copulas C and $\hat{C}$. For all $u, v \in [0, 1]$, we have

$$\hat{C}(1 - u, 1 - v) = 1 - u - v + C(u, v) \ . \tag{2.7}$$

Last but not least, we now introduce a version of Sklar's theorem for conditional distributions (Patton, 2006).

Theorem 2.2.3 (**Sklar's theorem, conditional version).** *Let X, Y, W be univariate r.v.'s and let $\mathcal{W}$ be the support of W. For every $w \in \mathcal{W}$, let $F_{X|W}(*|w)$ and $F_{Y|W}(*|w)$ be the CDFs of the conditional margins $[X|W = w]$ and $[Y|W = w]$. Also, let $F_{XY|W}(*, *|w)$ be the joint conditional CDF of $[(X, Y)|W = w]$. If $F_{X|W}(*|w)$ and $F_{Y|W}(*|w)$ are continuous, then there exists a unique 2-copula $C(*, *|w)$ such that*

$$F_{XY|W}(x, y|w) = C(F_{X|W}(x|w), F_{Y|W}(y|w)|w) \ , \tag{2.8}$$

for all $x, y \in \mathbb{R}$.

According to Patton:

> "It is the converse of Sklar's theorem that is the most interesting for multivariate density modelling."
>
> (Patton, 2006)

We next introduce the converse of Theorem 2.2.3.

Theorem 2.2.4 (**Converse of Theorem** 2.2.3**).** *Using the notation in Theorem* 2.2.3*, if $\{C(*, *|w)\}_{w \in \mathcal{W}}$ is a family of conditional copulas that is measurable in w, then $F_{XY|W}(*, *|w)$ defined in (2.8) is a conditional bivariate CDF with conditional margin CDFs $F_{X|W}(*|w)$ and $F_{Y|W}(*|w)$.*

2.3 Dependence measures

In Section 2.2 we saw how copulas naturally appear as linking functions between univariate margins. In this section we will look at some important dependence measures that happen to exclusively depend on copulas, reinforcing their pivotal role. In Nelsen's words:

> "... it is the copula which captures the 'non-parametric', 'distribution-free' or 'scale-invariant' nature of the association between random variables."
>
> (Nelsen, 2003)

2.3.1 Rank correlation

Given two r.v.'s X and Y, remember that the linear correlation coefficient between X and Y, also known as Pearson product-moment correlation coefficient, is defined as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \ , \tag{2.9}$$

where σ_X and σ_Y represent the standard deviations of X and Y, respectively, and $\mathrm{Cov}(X,Y)$ represents the covariance between X and Y. We always have $-1 \leq \rho(X,Y) \leq 1$.

The linear correlation (2.9) suffers from the following drawbacks:

1. $\rho(X,Y)$ is undefined if either $\mathbb{E}(X^2)$ or $\mathbb{E}(Y^2)$ is infinite.

2. If α and β are non-linear strictly increasing functions, we generally have $\rho(\alpha(X),\beta(Y)) \neq \rho(X,Y)$.

3. Given two univariate CDFs F and G and some $\lambda > 0$, we cannot generally expect to find a bivariate r.v. (X,Y) such that $X \sim F$, $Y \sim G$ and $\rho(X,Y) = \lambda$.

The rank correlation measures we will explore next overcome these drawbacks of the well-known linear correlation coefficient.

Concordance

The following function plays an important role to understand the rank correlation measures in this section (Nelsen, 2003).

Definition 2.3.1 (Concordance function). Let $(X_1,Y_1),(X_2,Y_2)$ be independent bivariate r.v.'s such that X_1 and Y_1 follow the same distributions as X_2 and Y_2, respectively. The concordance between (X_1,Y_1) and (X_2,Y_2) is defined as

$$\begin{aligned}\kappa_{(X_1,Y_1),(X_2,Y_2)} &= \underbrace{\mathbb{P}\left((X_1-X_2)(Y_1-Y_2)>0\right)}_{\text{concordance probability}} - \underbrace{\mathbb{P}\left((X_1-X_2)(Y_1-Y_2)<0\right)}_{\text{discordance probability}} \\ &= \mathbb{E}\left[\mathrm{sign}(X_1-X_2)(Y_1-Y_2)\right] \ .\end{aligned}$$

Interestingly, the concordance function exclusively depends on copulas.

Proposition 2.3.1. *The concordance $\kappa_{(X_1,Y_1),(X_2,Y_2)}$ exclusively depends on C_1 and C_2: the copulas of (X_1,Y_1) and (X_2,Y_2), respectively. Namely,*

$$\kappa_{(X_1,Y_1),(X_2,Y_2)} = 4\int_{[0,1]^2} C_1 \ dC_2 - 1 = 4\int_{[0,1]^2} C_2 \ dC_1 - 1 \ .$$

Consequently, we shall equivalently denote $\kappa_{(X_1,Y_1),(X_2,Y_2)}$ by $\kappa(C_1,C_2)$.

Spearman's rho

The following alternative to the linear correlation coefficient is actually based on it.

Definition 2.3.2 (Spearman's rho). Let X, Y be unidimensional continuous r.v.'s with CDFs F_X and F_Y, respectively. We define the Spearman's rho between X and Y as

$$\varrho(X,Y) = \rho(F_X(X), F_Y(Y)) \ .$$

Spearman's rho is always well-defined, since both $F_X(X)$ and $F_Y(Y)$ are uniformly distributed in $[0,1]$, according to Proposition C.0.14. Moreover, according to the following result, Spearman's rho only depends on the copula linking X and Y.

Proposition 2.3.2. *Spearman's rho only depends on the copula between X and Y. Namely, we have*

$$\varrho(X,Y) = 3\kappa(C_{X,Y}, C^{\perp}) \ .$$

Therefore, we can safely write $\varrho(C_{X,Y}) \equiv \varrho(X,Y)$.

Kendall's tau

Another alternative to the linear correlation coefficient is Kendall's tau.

Definition 2.3.3 (Kendall's tau). Let X, Y be unidimensional continuous r.v.'s. We define the Kendall's tau between X and Y as

$$\tau(X,Y) = \kappa_{(X,Y),(\bar{X},\bar{Y})} \ ,$$

where $(\bar{X}, \bar{Y})$ is an independent *copy* of (X, Y).

Since the concordance $\kappa_{(X,Y),(\bar{X},\bar{Y})}$ only depends on the copula $C_{X,Y}$ between X and Y (because $C_{\bar{X},\bar{Y}} = C_{X,Y}$), bearing in mind Proposition 2.3.1, we can safely write $\tau(C_{X,Y}) \equiv \tau(X,Y)$.

Kendall's tau is related to the Kendall distribution (Definition 2.1.5) of the copula $C_{X,Y}$,

Proposition 2.3.3. *Let X, Y be unidimensional continuous r.v.'s. We have*

$$\begin{aligned} \tau(X,Y) &= 4 \cdot \mathbb{E}\left(C_{X,Y}(F_X(X), F_Y(Y))\right) - 1 \\ &= 4 \cdot \mathbb{E}\left(F_{X,Y}(X,Y)\right) - 1 \end{aligned} \ .$$

Some important properties

Since both Spearman's rho and Kendall's tau exclusively depend on copulas and since copulas remain invariant under strictly increasing transformations α and β (Proposition 2.2.1), we have $\delta(\alpha(X), \beta(Y)) = \delta(X,Y)$ for both $\delta = \varrho$ and $\delta = \tau$. Other properties are summarized in the following proposition.

Proposition 2.3.4. *Let X, Y be unidimensional continuous r.v.'s with copula $C_{X,Y}$. Let δ represent either Spearman's rho ϱ or Kendall's tau τ. We have:*

1. *$\delta(X,Y) = 1$ if and only if $C_{X,Y} = C^{+}$.*
2. *$\delta(X,Y) = -1$ if and only if $C_{X,Y} = C^{-}$.*
3. *$\delta(X,Y) = 0$ if $C_{X,Y} = C^{\perp}$.*

2.3.2 Tail dependence

Tail dependence applies when analysing extreme events in both the upper and lower tail.

Definition 2.3.4 (Lower tail dependence). Let X, Y be unidimensional continuous r.v.'s. X and Y are said to be (asymptotically) dependent in the lower tail if the following limit exists and is positive:

$$\begin{aligned}\lambda_L(X,Y) &= \lim_{\alpha\to 0^+} \mathbb{P}\left(Y \le F_Y^{\leftarrow}(\alpha) | X \le F_X^{\leftarrow}(\alpha)\right) \\ &= \lim_{\alpha\to 0^+} \mathbb{P}\left(X \le F_X^{\leftarrow}(\alpha) | Y \le F_Y^{\leftarrow}(\alpha)\right)\end{aligned} . \tag{2.10}$$

If $\lambda_L(X,Y)$ exists but is zero, X and Y are said to be (asymptotically) independent in the lower tail.

Definition 2.3.5 (Upper tail dependence). Let X, Y be unidimensional continuous r.v.'s. X and Y are said to be (asymptotically) dependent in the upper tail if the following limit exists and is positive:

$$\begin{aligned}\lambda_U(X,Y) &= \lim_{\alpha\to 1^-} \mathbb{P}\left(Y \ge F_Y^{\leftarrow}(\alpha) | X \ge F_X^{\leftarrow}(\alpha)\right) \\ &= \lim_{\alpha\to 1^-} \mathbb{P}\left(X \ge F_X^{\leftarrow}(\alpha) | Y \ge F_Y^{\leftarrow}(\alpha)\right)\end{aligned} . \tag{2.11}$$

If $\lambda_U(X,Y)$ exists but is zero, X and Y are said to be (asymptotically) independent in the upper tail.

The limits (2.10) and (2.11) are known as the lower and upper tail indices, respectively. Interestingly, they both exclusively depend on copulas.

Proposition 2.3.5. *Let C be the copula between X and Y in both Definition* 2.3.4 *and Definition* 2.3.5*. We have that*

$$\lambda_L(X,Y) \equiv \lambda_L(C) = \lim_{u\to 0^+} \frac{C(u,u)}{u} ,$$

and

$$\lambda_U(X,Y) \equiv \lambda_U(C) = \lim_{u\to 1^-} \frac{1-2u+C(u,u)}{1-u} .$$

Remark 2.3.1. Note that $\lambda_U(C) = \lambda_L(\hat{C})$. Using (2.7):

$$\lim_{u\to 1^-} \frac{1-2u+C(u,u)}{1-u} = \lim_{u\to 1^-} \frac{\hat{C}(1-u,1-u)}{1-u} = \lim_{u\to 0^+} \frac{\hat{C}(u,u)}{u} .$$

2.4 Copula families

In the present section we will introduce some important examples of copulas that are actually used in real applications.

2.4.1 Elliptical copulas

In general, copulas arising from known distributions via Sklar's theorem are called *implicit*. An important family of implicit copulas is that of elliptical copulas. They are named after the family of elliptical distributions, from which they derive by means of Sklar's Theorem 2.2.1.

Elliptical distributions

Let us first remind some concepts about elliptical distributions (Cambanis et al., 1981), a *meta*-family that generalizes household distributions like the multivariate Gaussian and the Student's t.

Definition 2.4.1 (**Elliptically contoured CDF**). An n-variate CDF F is said to be elliptically contoured with parameters $(\boldsymbol{\mu}, \Sigma, \phi)$, represented by $F \in E_n(\boldsymbol{\mu}, \Sigma, \phi)$, if its Fourier-Stieltjes transform

$$\mathcal{F}\{F\}(\mathbf{t}) = \int_{\mathbb{R}^n} e^{i\mathbf{t}^T\mathbf{x}}\, dF(\mathbf{x}) \ ,$$

satisfies

$$\mathcal{F}\{F\}(\mathbf{t}) = e^{i\mathbf{t}^T\boldsymbol{\mu}} \cdot \phi\left(\mathbf{t}^T\Sigma\mathbf{t}\right) \ ,$$

for some:

- $\boldsymbol{\mu} \in \mathbb{R}^n$.
- $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$, symmetric and positive semi-definite[5].
- $\phi : [0, \infty) \to \mathbb{R}$.

The function ϕ is known as the distribution generator.

The name 'elliptical' is due to the fact that, if F is absolutely continuous, then its density is given by

$$f(\mathbf{x}) = (\det(\Sigma))^{-1/2}\, g\left((\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \ ,$$

for some non-negative function g of one real variable. Hence, the level sets of f are ellipsoids in $\mathbb{R}^n$ (Embrechts et al., 2003).

Definition 2.4.1 admits an equivalent interpretation in terms of r.v.'s.

Definition 2.4.2 (**Elliptical r.v.**). Let $(\boldsymbol{\mu}, \Sigma, \phi)$ as in Definition 2.4.1. An n-dimensional r.v. $\mathbb{X}$ is said to be elliptical with parameters $(\boldsymbol{\mu}, \Sigma, \phi)$, represented by $\mathbb{X} \in E_n(\boldsymbol{\mu}, \Sigma, \phi)$, if $F_{\mathbb{X}}$, the CDF of $\mathbb{X}$, is elliptical with parameters $(\boldsymbol{\mu}, \Sigma, \phi)$ or, equivalently, if the characteristic function of $\mathbb{X}$,

$$\varphi_{\mathbb{X}}(\mathbf{t}) = \mathbb{E}\left(e^{i\mathbf{t}^T\mathbb{X}}\right) \ ,$$

satisfies

$$\varphi_{\mathbb{X}}(\mathbf{t}) = e^{i\mathbf{t}^T\boldsymbol{\mu}} \cdot \phi\left(\mathbf{t}^T\Sigma\mathbf{t}\right) \ .$$

Note that the parameters that define an elliptical distribution are not unique.

Remark 2.4.1. If $F \in E_n(\boldsymbol{\mu}, \Sigma, \phi)$ and $\lambda > 0$, then $F \in E_n(\boldsymbol{\mu}, \lambda\Sigma, \phi(*/\lambda))$.

Definition 2.4.2 is equivalent to the following handy representation (Cambanis et al., 1981, Embrechts et al., 2003).

[5] $\mathbf{t}^T\Sigma\mathbf{t} \geq 0$, for all $\mathbf{t} \in \mathbb{R}^n$.

Theorem 2.4.1 (**Stochastic representation of elliptical r.v.'s**). *An n-dimensional r.v. $\mathbb{X}$ satisfies $\mathbb{X} \in E_n(\boldsymbol{\mu}, \Sigma, \phi)$ with $\operatorname{rank}(\Sigma) = k$ if and only if the following equality in distribution holds:*

$$\mathbb{X} \overset{d}{=} \boldsymbol{\mu} + RA\mathbb{U} \ ,$$

where

- *R is a unidimensional r.v. whose CDF F_R satisfies*

$$\phi(x) = \int_0^\infty \varphi_{\mathbb{U}}(r^2 x) \ dF_R(r) \ , \tag{2.12}$$

  where $\varphi_{\mathbb{U}}$ represents the characteristic function of $\mathbb{U}$.

- *$A \in \mathcal{M}_{n\times k}(\mathbb{R})$ verifies $AA^T = \Sigma$.*

- *$\mathbb{U}$ is a k-variate r.v. uniformly distributed over the unit hypersphere $\{\mathbf{u} \in \mathbb{R}^k : \mathbf{u}^T\mathbf{u} = 1\}$ and independent of R.*

Looking at (2.4.1), it does not seem too far-fetched and it can actually be demonstrated that the univariate margins of elliptical distributions also belong to the elliptical family. Copulas arising from elliptical distributions thus provide an opportunity to build multivariate distributions, via (Proposition 2.2.2), where the univariate margins can be arbitrarily chosen: they are called meta-elliptical distributions. Notwithstanding, elliptical copulas are always radially symmetric (Definition 2.2.1), as pointed out by (McNeil et al., 2005). Hence, working with elliptical copulas is fairly restrictive.[6]

Two major elliptical copulas

We next present two major examples of copulas arising from elliptical distributions.

Example 2.4.1. The n-variate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ is elliptical. If $\mathbb{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, we have

$$\varphi_{\mathbb{X}}(\mathbf{t}) = \exp\left(i\mathbf{t}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right) \ ,$$

and thus $\mathbb{X} \in E_n(\boldsymbol{\mu}, \Sigma, \phi)$, for $\phi(x) = e^{-x/2}$.

From Example 2.4.1 and using (2.6) we get the Gaussian copula.

Definition 2.4.3 (**Gaussian copula**). Let $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$ be a covariance matrix with $\Sigma_{ii} = \sigma_i^2 = 1$, for all $i = 1, 2, \ldots, n$. We define the n-variate Gaussian copula with parameter matrix Σ as

$$C_\Sigma^{Ga}(u_1, u_2, \ldots, u_n) = \Phi_\Sigma\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right) \ , \tag{2.13}$$

where Φ and Φ_Σ are the CDFs of $\mathcal{N}(0,1)$ and $\mathcal{N}_n(\mathbf{0}_n, \Sigma)$, respectively.

[6] For instance, according to Proposition 2.3.5, upper and lower tail indices are necessarily identical for radially-symmetric copulas.

The parameter matrix Σ in (2.4.3) is actually a correlation matrix. Interestingly, all multivariate Gaussian distributions sharing a certain correlation matrix produce the same copula (2.13), regardless of the means and variances of their univariate margins. Indeed, this is a consequence of Proposition 2.2.1, since the standardizing transformations

$$\alpha_i : X_i \longmapsto \frac{X_i - \boldsymbol{\mu}_i}{\sigma_i} \ ,$$

which ensure zero means and unit standard deviations, are strictly increasing.

The second example derives from the Student's t distribution.

Example 2.4.2. The n-variate Student's t distribution is elliptical. Remember that $\mathbb{X}$ follows an n-variate Student's t distribution with parameters ν, $\boldsymbol{\mu}$ and Σ, represented by $\mathbb{X} \sim t_n(\nu, \boldsymbol{\mu}, \Sigma)$, if

$$\mathbb{X} \overset{d}{=} \boldsymbol{\mu} + \frac{\sqrt{\nu}}{\sqrt{S}}\mathbf{Z} \ , \tag{2.14}$$

where

- $\nu \in \mathbb{R}$, $\nu > 0$.
- $\boldsymbol{\mu} \in \mathbb{R}^n$.
- $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$ is symmetric and positive semi-definite.
- $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}_n, \Sigma)$.
- S follows a chi-squared distribution with ν degrees of freedom, $S \sim \chi^2_\nu$, and is independent of $\mathbf{Z}$.

Again, using (2.6), we get the Student's copula.

Definition 2.4.4 (Student's copula). Let $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$ be a scale matrix with $\Sigma_{ii} = \sigma_i^2 = 1$, for all $i = 1, 2, \ldots, n$. We define the n-variate Student's copula with parameter matrix Σ as

$$C^t_{\nu,\Sigma}(u_1, u_2, \ldots, u_n) = t^n_{\nu,\Sigma}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_n)) \ ,$$

where t_ν and $t^n_{\nu,\Sigma}$ are the CDFs of $t(\nu, 0, 1)$ and $t_n(\nu, \mathbf{0}_n, \Sigma)$, respectively.

Just like in Definition 2.4.3, Student's copula can be fully described in terms of a scale matrix Σ with ones in the main diagonal. Both the Gaussian copula and Student's copula are *parametric* models, much like the original distributions they derive from. For the n-variate Gaussian copula we have to estimate (using, say, maximum likelihood estimation) $n(n-1)/2$ correlation coefficients, whereas for Student's copula there is an additional parameter: the degrees of freedom ν.

2.4.2 Archimedean copulas

The elliptical copulas reviewed in the previous section fall into the implicit copulas category, since they arise from previously known multivariate distributions. By contrast, the copulas we will next introduce, which are the cornerstone of the modelling methods proposed in this work, have an explicit expression in terms of a functional parameter.

Definition 2.4.5 (Archimedean copula). Let $\phi : [0,1] \to [0,\infty]$ be a strictly decreasing, convex function satisfying $\phi(1) = 0$. We define the Archimedean copula with generator ϕ as

$$C_\phi(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \ , \tag{2.15}$$

where

$$\phi^{[-1]}(x) = \begin{cases} \phi^{-1}(x), & \text{if } 0 \le x < \phi(0) \\ 0, & \text{otherwise} \end{cases} \ . \tag{2.16}$$

We shall denote the set of Archimedean generators by $\mathbf{\Phi}$.

Archimedean copulas rely on a single functional parameter with relatively simple axioms. As compared with the *crude* Definition 2.1.1 and leaving out computational considerations, Definition 2.4.5 seems an easy procedure to obtain copulas: one can easily *draw* an Archimedean generator (or at least a bounded one). One can show that the convexity requirement for the generator ϕ is necessary (Embrechts et al., 2003). In other words, we cannot find a copula of the form (2.15) for a non-convex generator ϕ.

The definition of an Archimedean copula in terms of the functional parameter is not unique: infinitely many generators produce the same Archimedean copula.

Proposition 2.4.1. *Let $\phi, \tilde{\phi} \in \mathbf{\Phi}$. We have that $C_\phi = C_{\tilde{\phi}}$ if and only there exists $\lambda > 0$ such that $\tilde{\phi}(x) = \lambda\phi(x)$, for all $x \in [0,1]$.*

Archimedean copulas are named after the Archimedean property they satisfy, according to the following result (Nelsen, 2006).

Proposition 2.4.2. *Let C_ϕ be an Archimedean copula with generator ϕ. Let $\mathbb{I} = [0,1]$ be endowed with the binary operator $\star : \mathbb{I} \times \mathbb{I} \to \mathbb{I}$ given by*

$$u \star v = C_\phi(u,v) \ .$$

It holds that $(\mathbb{I}, \star, \le)$ is a commutative totally ordered monoid. Moreover, considering the exponentiation in $(\mathbb{I}, \star)$, which maps $u \in \mathbb{I}$ to its k-th power, for $k \in \mathbb{Z}_+$,

$$u^{k)} = \underbrace{u \star u \star \ldots \star u}_{k \text{ times}} = \begin{cases} u \star u^{k-1)}, & \text{if } k \ge 1 \\ 1, & \text{if } k = 0 \end{cases} \ ,$$

$(\mathbb{I}, \star, \le)$ satisfies the following version of the Archimedean property[7]: for every $u, v \in \mathbb{I}$, if $u < 1$ and $v < 1$, there exists $n \in \mathbb{N}$ such that $u^{n)} < v$.

The Archimedean family is rather extensive, gathering a wide variety of dependencies. In fact, two fundamental copulas we have seen are Archimedean.

Example 2.4.3. The minimal copula C^- is Archimedean with generator $\phi(x) = 1 - x$.

Example 2.4.4. The bivariate independence copula $C^\perp$ is Archimedean with generator $\phi(x) = -\log x$.

[7] The Archimdean property in $(\mathbb{R}, +, \le)$ can be stated as: for every $x, y \in \mathbb{R}$, if $x > 0$ and $y > 0$, there exists $n \in \mathbb{N}$ such that $n \times x > y$.

Family	**Generator** $\phi(x)$	**Parameter range**
Gumbel	$(-\log x)^\theta$	$\theta \in [1, \infty)$
Clayton	$\frac{1}{\theta}(x^{-\theta} - 1)$	$\theta \in [-1, 0) \cup (0, \infty)$
Frank	$-\log\left(\frac{e^{-\theta x}-1}{e^{-\theta}-1}\right)$	$\theta \in (-\infty, 0) \cup (0, \infty)$

Table 2.1: Some important Archimedean families.

Table 2.1 summarizes three of the foremost and most widespread Archimedean (sub)families in the literature. In all cases, the generator is parametrized by a certain θ. For this reason, these Archimedean copulas can be seen as parametric, even though they all derive from the more general construction given in Definition 2.4.5.

The convexity of ϕ implies some smoothness properties worth mentioning.

Remark 2.4.2. An Archimedean generator ϕ is continuous on $(0, 1]$ and differentiable at all points, except for a countable subset of $(0, 1)$. Nonetheless, left and right derivatives always exist.

Archimedean copulas are usually endowed with the following extra convenient hypothesis.

Definition 2.4.6 (**Strict Archimedean copula).** An Archimedean copula and its generator ϕ are said to be strict, represented by $\phi \in \mathbf{\Phi}_\infty$, if $\phi(0) = \infty$.

The strictness property $\phi(0) = \infty$ reduces (2.15) to simply $C(u, v) = \phi^{-1}(\phi(u)+\phi(v))$. It turns out that the definition of a strict Archimedean copula can indistinctly be stated in terms of either ϕ or ϕ^{-1} as generator. The key point in using strict Archimedean copulas is contained in the following proposition.

Proposition 2.4.3. *Let C_ϕ be an Archimedean copula with generator ϕ. The following statements are equivalent:*

1. *ϕ is strict, i.e., $\phi(0) = \infty$.*
2. *For all $(u, v) \in (0, 1]^2$, we have $C_\phi(u, v) > 0$.*

Proof. If ϕ is strict, then $\phi^{[-1]}(t) = \phi^{-1}(t) > 0$ for all $t \in [0, \infty)$. Taking $u \neq 0$ and $v \neq 0$, we have $\phi(u) + \phi(v) < \infty$, so $C(u, v) > 0$. Conversely, if $\phi(0) < \infty$, since ϕ is continuous, we can find $u > 0$ such that $2\phi(u) > \phi(0)$. Therefore, $C_\phi(u, u) = 0$, which means $\mu_{C_\phi}([0, u]^2) = 0$. □

Proposition 2.4.3 means that the support of a bivariate distribution whose copula is Archimedean is not the product of its margin supports unless $\phi(0) = \infty$. This feature of non-strict Archimedean copulas makes them unsuitable for most applications.

Archimedean copulas and Kendall distribution

There is an interesting connection between Archimedean copulas and the Kendall distribution (Definition 2.1.5). To begin with, Archimedean copulas have a simple expression for (2.5).

Proposition 2.4.4 (Kendall distribution for Archimedean copulas). *Let C_ϕ be an Archimedean copula with generator ϕ. We have that*

$$K_{C_\phi}(x) = x - \frac{\phi(x)}{\phi'(x^+)} \ . \tag{2.17}$$

Given a function $K : (0,1) \to \mathbb{R}$ playing the role of K_{C_ϕ}, equation (2.17) is actually a first order ordinary differential equation for an unknown Archimedean generator $\phi \in \mathcal{C}^1(0,1)$:

$$\frac{\phi'(x)}{\phi(x)} = \frac{1}{x - K(x)} \ , \tag{2.18}$$

whose general solution is

$$\phi(x) = \phi(x_0) \cdot \exp\left(\int_{x_0}^{x} \frac{dt}{t - K(t)} \right) \ . \tag{2.19}$$

Observe that, fixed $x_0 \in (0,1)$, all solutions (2.19) differ by a positive constant $\phi(x_0)$. Nonetheless, for Proposition 2.4.1, all of them produce the same Archimedean copula. Therefore, there can only exist one Archimedean copula given K. (Genest and Rivest, 1993) gives a necessary and sufficient condition for (2.19) to be an Archimedean generator.

Theorem 2.4.2. *Let K be a univariate CDF. The function ϕ in (2.19) is a solution for the problem (2.18), i.e., ϕ generates an Archimedean copula, if and only if $K(x^-) = \lim_{t \to x^-} K(t) > x$, for all $x \in (0,1)$.*

Dependence measures revisited

Let us now examine the dependence measures studied in Section 2.3 in the context of Archimedean copulas. We can find a simple expression (Nelsen, 2003) for Kendall's tau (Definition 2.3.3) by combining Proposition 2.3.3 and equation (2.17).

Proposition 2.4.5. *Let C_ϕ be an Archimedean copula with generator ϕ. We have*

$$\begin{aligned} \tau(C_\phi) &= 4 \int_0^1 x \ dK_{C_\phi}(x) - 1 \\ &= 3 - 4 \int_0^1 K_{C_\phi}(x) \ dx \ . \\ &= 1 + 4 \int_0^1 \frac{\phi(x)}{\phi'(x^+)} \ dx \end{aligned}$$

As regards tail dependence, there exist explicit expressions for both tail indices in terms of $(\phi^{-1})'$ (Embrechts et al., 2003), but they generally only hold for a special class of Archimedean copulas that are out of the scope of this work: Laplace Transform (LT) Archimedean copulas. Even though all copula families in Table 2.1 belong to the LT class, the approach we will follow does not rely on the same hypotheses. A convenient additional alternative hypothesis for Archimedean generators when modelling tail dependence is regular variation.[8]

[8] See Appendix B for further details.

Family	τ	λ_U	λ_L
Gumbel	$1 - 1/\theta$	$2 - 2^{1/\theta}$	0
Clayton	$\theta/(\theta + 2)$	0	$\begin{cases} 2^{-1/\theta}, & \text{if } \theta > 0 \\ 0, & \text{if } \theta < 0 \end{cases}$
Frank	$1 - 4\theta^{-1}(1 - D_1(\theta))$	0	0

Table 2.2: Dependence measures for the Archimedean parametric families in Table 2.1. D_1 is the Debye function $D_1(\theta) = \theta^{-1} \int_0^\theta \frac{t}{e^t - 1}\, dt$.

Proposition 2.4.6. *Let C_ϕ be an Archimedean copula with generator ϕ.*

1. *If $\phi \in \mathcal{R}_{0^+,\alpha}$, for some $-\infty < \alpha \leq 0$, then $\lambda_L(C_\phi) = 2^{-1/|\alpha|}$.*[9]

2. *If $\phi \in \mathcal{R}_{1^-,\alpha}$, for some $1 \leq \alpha < \infty$, then $\lambda_U(C_\phi) = 2 - 2^{1/\alpha}$.*

We know that regular variation for continuously differentiable monotone functions is alternatively characterized by the existence of certain limits, according to Proposition B.0.8, and of course Archimedean generators are strictly decreasing. Therefore, $\phi \in \mathcal{R}_{0^+,\alpha}$ is equivalent to

$$\lim_{x \to 0^+} \frac{x\phi'(x)}{\phi(x)} = \alpha \ , \tag{2.20}$$

whereas $\phi \in \mathcal{R}_{1^-,\alpha}$ is equivalent to

$$\lim_{x \to 1^-} \frac{(1-x)\phi'(x)}{\phi(x)} = -\alpha \ . \tag{2.21}$$

Note that the convexity of ϕ ensures $(1-x)\phi'(x) \leq -\phi(x)$ and necessarily $\alpha \geq 1$ in (2.21). Charpentier and Segers state for both (2.20) and (2.21):

> "The limit indeed exists for virtually every known parametric model."
>
> (Charpentier and Segers, 2009)

Therefore, according to Charpentier and Segers, regular variation "is not a very restrictive assumption". In fact, using (2.17), limits (2.20) and (2.21) are equivalent to the less *intimidating* limits

$$K'_{C_\phi}(0^+) = \lim_{x \to 0^+} \frac{K_{C_\phi}(x)}{x} = 1 + \frac{1}{|\alpha|}$$

and

$$K'_{C_\phi}(1^-) = \lim_{x \to 1^-} \frac{1 - K_{C_\phi}(x)}{1 - x} = 1 - \frac{1}{\alpha} \ ,$$

respectively.

[9] Using the convention $1/0 = \infty$.

2.5 Basic copula estimation methods

In the context of copula modelling, estimation refers to the process of obtaining a *suitable* copula representation of the underlying dependence between samples of different r.v.'s. Such a copula model can then be used for inference.

Copula estimation methods are either parametric or non-parametric. In parametric methods, the copula is assumed to have a particular functional form that depends on a finite, usually small number of parameters. We have already seen some examples of copula families that depend on one or more parameters: the Gaussian and Student's copula families and those summarized in Table 2.1. By contrast, in non-parametric methods one does not assume that the copula belongs to a specific parametric family, but rather directly *build* the model from available data.

In what follows, we shall focus on bivariate copulas for the sake of simplicity.

2.5.1 Parametric estimation

Suppose we want to estimate the joint CDF H of a bivariate r.v. (X, Y) from a sample $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$. Let F and G be the unknown univariate CDFs of X and Y, respectively. For Sklar's theorem, we know that $H(x, y) = C(F(x), G(y))$, for some bivariate copula C. Therefore, a reasonable strategy consists in conjecturing some parametric models $F_{\boldsymbol{\alpha}}$, $G_{\boldsymbol{\beta}}$ and $C_{\boldsymbol{\gamma}}$, depending on vector parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, to finally obtain a parametric model $H_{\boldsymbol{\theta}}$ that relies on a vector $\boldsymbol{\theta}$ gathering all the previous ones.[10]

Letting $f_{\boldsymbol{\alpha}}$, $g_{\boldsymbol{\beta}}$, $c_{\boldsymbol{\gamma}}$ and $h_{\boldsymbol{\theta}}$, be the densities of $F_{\boldsymbol{\alpha}}$, $G_{\boldsymbol{\beta}}$, $C_{\boldsymbol{\gamma}}$ and $H_{\boldsymbol{\theta}}$, respectively, the following equation holds:

$$h_{\boldsymbol{\theta}}(x, y) = c_{\boldsymbol{\gamma}}(F_{\boldsymbol{\alpha}}(x), G_{\boldsymbol{\beta}}(y)) \cdot f_{\boldsymbol{\alpha}}(x) \cdot g_{\boldsymbol{\beta}}(y) \ . \tag{2.22}$$

Now, we can estimate $\boldsymbol{\theta}$ from data $\mathcal{D}$ by maximizing the likelihood function

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i=1}^{N} h_{\boldsymbol{\theta}}(X_i, Y_i) \ .$$

Model (2.22) is actually a parametric model of the density of H, where parameters for both the univariate margins and the copula are estimated all at once. In copula theory, though, it is quite common to separate both estimation types in different stages:

1. Optimize the parameters for the univariate margins:

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \left\{ \prod_{i=1}^{N} f_{\boldsymbol{\alpha}}(X_i) \right\} \ , \ \hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \left\{ \prod_{i=1}^{N} g_{\boldsymbol{\beta}}(Y_i) \right\} \ .$$

2. Using the previous estimations for the univariate margins, estimate the parameters of the copula:

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \left\{ \prod_{i=1}^{N} c_{\boldsymbol{\gamma}}(F_{\hat{\boldsymbol{\alpha}}}(X_i), G_{\hat{\boldsymbol{\beta}}}(Y_i)) \right\} \ .$$

[10] Consider the parameter concatenation $\boldsymbol{\theta} = (\boldsymbol{\alpha}; \boldsymbol{\beta}; \boldsymbol{\gamma})$, for instance.

3. Finally, take $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}; \hat{\boldsymbol{\beta}}; \hat{\boldsymbol{\gamma}})$.

According to (Bouyé et al., 2000), the former approach is called *exact maximum likelihood method* (EML), whereas the latter is known as the *inference functions for margins* method (IFM).

There is another parametric estimation approach yet which generalizes IFM. Note that the univariate margins F and G need not be modelled in a parametric manner. Instead, consider generic estimates $\hat{F}$ and $\hat{G}$, obtained by any valid means, and build the pseudo-sample $\mathcal{D}' = \{(U_i, V_i)\}_{i=1}^N$, where $U_i = \hat{F}(X_i)$ and $V_i = \hat{G}(Y_i)$, for $i = 1, 2, \ldots, N$. Finally, find the $\boldsymbol{\gamma}$ that maximizes

$$\mathcal{L}(\boldsymbol{\gamma}|\mathcal{D}') = \prod_{i=1}^{N} c_{\boldsymbol{\gamma}}(U_i, V_i) \ . \tag{2.23}$$

Observe that the pseudo-sample $\mathcal{D}'$ is an approximation to a sample arising from the bivariate r.v. $(U, V) \equiv (F(X), G(Y))$. Therefore, maximizing (2.23) is an attempt to estimate the parameters of c as the density of a bivariate r.v. (with uniform margins), not just the parameters of a linking function.

Some researchers, namely Mikosch, argue against the (2.23) approach:

> "The marginal distributions and the copula of a multivariate distribution are inextricably linked. The main selling point of the copula technology — separation of the copula (dependence function) from the marginal distributions — leads to a biased view of stochastic dependence, in particular when one fits a model to the data."
>
> (Mikosch, 2006)

Nonetheless, there are actually serious reasons to be interested in a two-phase estimation procedure. (Genest and Rémillard, 2006) demonstrates, using Proposition 2.1.1, that the approximation error for a multivariate distribution is bounded from above by the sums of the separate errors made in both the copula and the margins estimation:

> "Indeed, basic facts from copula theory imply that a good fit of the joint distribution function necessarily results from good fits of the copula and the margins taken separately, especially in low dimension."
>
> (Genest and Rémillard, 2006)

In the bivariate case: letting $\hat{C}$ be the estimate of a copula C arising from (2.23) and letting $H(x, y) = C(F(x), G(y))$ and $\hat{H}(x, y) = \hat{C}(\hat{F}(x), \hat{G}(y))$, we have

$$\|H - \hat{H}\|_\infty \leq \|C - \hat{C}\|_\infty + \|F - \hat{F}\|_\infty + \|G - \hat{G}\|_\infty \ ,$$

where $\|*\|_\infty$ is the uniform norm. Moreover, Genest and Rémillard remind that, if $\breve{H}$ is any other estimate of H, "things can go terribly wrong" for a copula deduced from $\breve{H}$ if "an inappropriate choice of margins is made".

All the above approaches are likelihood-based, but this is not the only parametric method. For instance, (Genest and Rivest, 1993) proposes to estimate the parameter θ for the families in Table 2.2 by solving for θ in $\hat{\tau} = \tau(\theta)$, where $\hat{\tau}$ is a sample estimate of the Kendall's tau. Therefore, this could be considered a method-of-moments estimation based on Kendall's tau. Similar approaches can be envisioned for more than one parameter.

2.5.2 Non-parametric estimation

There are lots of different ways to estimate a copula in a non-parametric manner, but many of them are essentially smoothed versions of the approach we present here (Charpentier et al., 2006). Since a copula C is, in particular, a CDF, we may approximate C with its empirical estimate

$$\tilde{C}(u,v) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[U_i,\infty)\times[V_i,\infty)}(u,v) \ , \tag{2.24}$$

where the U_i's and V_i's are taken from the pseudo-sample $\mathcal{D}'$ in the previous section.

Even though the empirical copula (2.24) approximates C, $\tilde{C}$ is *not* a copula. To see this, note that every copula is continuous, but (2.24) is not.

2.6 Vines

The vine construction can be used to build a 3-copula using 2-copulas as building blocks. *Vines* (Bedford and Cooke, 2002) can be used also to build copulas in the more general n-variate case, but since we are exclusively dealing with copulas up to $n = 3$ in this work, we will not introduce some concepts relative to discrete structures (trees, vertices, edge, and so on) that are needed in the general definition. Hence, the following definition is not the original one in (Bedford and Cooke, 2002) or (Lopez-Paz et al., 2013), but a simplified version, which is sufficient for our purposes.

Definition 2.6.1 (Vine). A vine is a 3-tuple $\big(C_{12}, C_{23}, \{C_{13|2}(*,*|w)\}_{w\in[0,1]}\big)$, where C_{12} and C_{23} are 2-copulas and $\{C_{13|2}(*,*|w)\}_{w\in[0,1]}$ is a family of bivariate copulas that is measurable in $w \in [0,1]$.

Vines uniquely define 3-copulas, according to the following result.

Proposition 2.6.1. *Let $\mathcal{V} = \big(C_{12}, C_{23}, \{C_{13|2}(*,*|w)\}_{w\in[0,1]}\big)$ be a vine. If we define, for each $w \in [0,1]$,*

$$F_{\mathcal{V}}(x,y|w) = C_{13|2}(\partial_2 C_{12}(x,w), \partial_1 C_{23}(w,y)|w) \ , \tag{2.25}$$

then the vine $\mathcal{V}$ uniquely defines a 3-copula $C_{\mathcal{V}}$ by means of

$$C_{\mathcal{V}}(u_1,u_2,u_3) = \int_0^{u_2} F_{\mathcal{V}}(u_1,u_3|w)\ dw \ . \tag{2.26}$$

Proof. We just verify the 3-increasingness property in Definition 2.1.1. First, note that (2.25) is a conditional bivariate CDF, for Theorem 2.2.4 and Propo-

sition 2.1.2. Then, given $R = (a_1, a_2] \times (b_1, b_2] \times (c_1, c_2] \subset \mathbb{R}^3$,

$$\begin{aligned}
\mu_{C_\mathcal{V}}(R) &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}(-1)^{i+j+k} \cdot C_\mathcal{V}(a_i, b_j, c_k) \\
&= \sum_{i=1}^{2}\sum_{k=1}^{2}(-1)^{i+k} \cdot \int_{b_1}^{b_2} F_\mathcal{V}(a_i, c_k|w)\ dw \\
&= \int_{b_1}^{b_2} \sum_{i=1}^{2}\sum_{k=1}^{2}(-1)^{i+k} \cdot F_\mathcal{V}(a_i, c_k|w)\ dw \\
&= \int_{b_1}^{b_2} \mu_{F_\mathcal{V}(*,*|w)}\left((a_1, a_2] \times (c_1, c_2]\right)\ dw \geq 0
\end{aligned} ,$$

where we have used that $F_\mathcal{V}(*, *|w)$ is 2-increasing. □

Remark 2.6.1. Using the notation in Proposition 2.6.1, one can check that $C_\mathcal{V}(u_1, u_2, 1) = C_{12}(u_1, u_2)$ and $C_\mathcal{V}(1, u_2, u_3) = C_{23}(u_2, u_3)$, for all $u_1, u_2, u_3 \in [0, 1]$.

Vines not only provide a way of constructing trivariate copulas, but also a decomposition scheme for 3-copulas, for Theorem 2.2.3.

Proposition 2.6.2. *Let C be a 3-copula and:*

- *Let $C_{1,2}$ be the copula linking the first and second univariate margins of C, that is, $C_{1,2}(u_1, u_2) = C(u_1, u_2, 1)$.*
- *Let $C_{2,3}$ be the copula linking the second and third univariate margins of C, that is, $C_{2,3}(u_2, u_3) = C(1, u_2, u_3)$.*

Assume the partial derivatives $\partial_2 C_{1,2}(*, w)$ and $\partial_1 C_{2,3}(w, *)$ are continuous, for all $w \in [0, 1]$. There exists a vine $\mathcal{V}$ such that $C = C_\mathcal{V}$, namely $\mathcal{V} = \left(C_{1,2}, C_{2,3}, \{C_{1,3|2}(*, *|w)\}_{w\in[0,1]}\right)$, where $C_{1,3|2}(*, *|w)$ is the unique copula linking the conditional margins $\partial_2 C_{1,2}(*, w)$ and $\partial_1 C_{2,3}(w, *)$, for Theorem 2.2.3.

Proof. Let (U_1, U_2, U_3) be a r.v. with CDF C. We have:

$$\begin{aligned}
C(u_1, u_2, u_3) &= \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, U_3 \leq u_3) \\
&= \int_{\{U_2 \leq u_2\}} \mathbb{P}(U_1 \leq u_1, U_3 \leq u_3|U_2)(\omega)\ d\mathbb{P}(\omega) \\
&= \int_0^{u_2} \mathbb{P}(U_1 \leq u_1, U_3 \leq u_3|U_2 = w)\ dw \\
&= \int_0^{u_2} C_{1,3|2}(\partial_2 C_{1,2}(u_1, w), \partial_1 C_{2,3}(w, u_3)|w)\ dw \\
&= C_\mathcal{V}(u_1, u_2, u_3)\ ,
\end{aligned}$$

where we have used the probabilistic interpretation of the partial derivatives of a copula (Proposition 2.1.2). □

The assumption that the copula between $[U_1|U_2 = u_2]$ and $[U_3|U_2 = u_2]$ does not change with the covariate value $u_2 \in [0, 1]$ is lately known as the 'simplifying assumption' (Hernández-Lobato and Suárez, 2011). This is equivalent to considering $C_{13|2}(*, *|w) = C$, for a certain bivariate copula C and all $w \in [0, 1]$. Under the simplifying assumption, the resulting vine structure is commonly known as a 'pair-copula construction' (PCC). Despite the fact that PCCs can be sometimes 'misleading' (Acar et al., 2012), they offer some flexibility while preserving simplicity. When the covariate effect is hardly noticeable, PCCs can be useful, as pointed out in (Lambert, 2014).

Just like we can model a multivariate CDF by separately specifying a copula and several univariate margins, vines allow to obtain a 3-variate copula from two bivariate margin copulas and a family of conditional copulas. Similarly, one may attempt to estimate the former two components of a vine, namely the conditional copulas and the bivariate margins, independently. Despite Mikosch's claim against this kind of modular estimation, this is indeed the preferred way to use vines.

Chapter 3

Previous work

> – Es... una integral rigeliana –tartamudeó el estudiante–, referida a una distribución planetaria que indica la presencia de dos clases económicas principales en el planeta, o tal vez un Sector, además de una pauta emocional inestable.
> – ¿Y qué significa?
> – Representa el límite de tensión, ya que aquí tenemos –señaló, y de nuevo se movieron las ecuaciones– una serie convergente.
>
> *Segunda fundación*
> Isaac Asimov

The present chapter is devoted to the review of several copula modelling techniques. In Chapter 2 we saw that multivariate[1] copula modelling can be modularly broken down, via *vine* constructions, into two separate estimation stages that can be carried out independently of each other; hence the appeal of this approach. The first part of the chapter addresses the estimation of the bivariate copula margins in a *vine* setting, introducing procedures other than parametric and empirical approaches. The second part completes the *vine* construction by focusing on the estimation of conditional copula families.

3.1 Bivariate copula modelling

In the present section we will look at *semiparametric* bivariate copula modelling strategies. This name suggests that such a method combines some properties of parametric and non-parametric methods, which were reviewed in Chapter 2. Semiparametric methods attempt to overcome some of the drawbacks of both parametric and non-parametric ones, which are associated with the bias and variance components of the generalization error. According to Domingos:

> "Bias is a learner's tendency to consistently learn the same wrong thing. Variance is the tendency to learn random things irrespective of the real signal."
>
> (Domingos, 2012)

[1] d-copulas, for $d \geq 3$.

Roughly speaking, a model with strong assumptions, like a linear classifier (Domingos, 2012), is too simple or *rigid* to capture a complex structure, which affects the bias component. On the other hand, a model with few assumptions, like a decision tree (Domingos, 2012), is more flexible, but also more sensitive to changes in the training set, thus affecting the variance component. The bias and variance components of the generalization error are associated with two problems: underfitting and overfitting, respectively. In Domingos's words: "It's easy to avoid overfitting (variance) by falling into the opposite error of underfitting (bias)". Then, the relation of parametric and non-parametric methods with underfitting and overfitting, respectively, is expressed by Hernández-Lobato and Suárez in:

> "While parametric copulas often lack expressive capacity to capture the complex dependencies that are usually found in empirical data, non-parametric copulas can have poor generalization performance because of overfitting."
>
> (Hernández-Lobato and Suárez, 2011)

There does not seem to be a unanimous definition among researchers for the term 'semiparametric'. The definition that, in our opinion, best represents the essence of semiparametric models is presented in (Kosorok, 2008), where it is stated, letting Θ be the parameter space of the model, that "Semiparametric models are statistical models where Θ has one or more infinite-dimensional component.". In the context of copula modelling, an infinite-dimensional parameter space usually refers to a function space. By contrast, for instance, Vandenhende and Lambert hold the opinion that semiparametric means "highly parameterized".

The main difficulty in the modelling process is to satisfy the rather stringent conditions of a copula, while leaving some flexibility for data fitting. In Chapter 2 we introduced Archimedean copulas, which are formulated in terms of a functional parameter ϕ, commonly known as the Archimedean generator. This generator fulfils some specific requirements. Nonetheless, these are simpler to model than the copula itself.

The Archimedean copula families described in Section 2.4.2 depend on a single parameter θ. More flexible Archimedean copulas can be build if the generator ϕ is selcted among a sufficiently large collection of flexible and well-behaved functions. In this work we propose to use polynomial splines for this purpose. For a quick reference on this topic, see Appendix A.

To end this introduction, we must acknowledge that Archimedean copulas are not the only types of copulas that are expressed in terms of a functional parameter. Semiparametric approaches could also be envisioned for extreme-value copulas (Gudendorf and Segers, 2010) or a more recently introduced family, the Archimax copulas (Charpentier et al., 2014), which depend on two functional parameters and include both Archimedean and extreme-value copulas.

3.1.1 A piecewise linear generator model

In this section we introduce the work by Vandenhende and Lambert (2005), a simple procedure to add more flexibility to Archimedean copulas. By looking at this technique, we will discover some of the difficulties we can expect when

modelling copulas as well as some hints on the key features a good model should have. Also, apart from being an ingenious and efficient technique, it will help us to grasp another one we shall present later on.

The essence of Vandenhende and Lambert's approach is contained in the following proposition.

Proposition 3.1.1. *If $\phi \in \mathbf{\Phi}$ and $\theta \geq 1$, then $x \mapsto (\phi(x))^\theta \in \mathbf{\Phi}$.*

Vandenhende and Lambert's original formulation, which we present next, apparently has nothing to do with Proposition 3.1.1, but the connection will soon become clear.

Proposition 3.1.2. *Let $\{t_i\}_{i=1}^n \subset \mathbb{R}$ satisfying $t_1 < t_2 < \ldots < t_n$. Consider the function $p : \mathbb{R} \to \mathbb{R}$ given by*

$$p(x) = \sum_{i=0}^{n} (\mu_i + \theta_i x) \mathbb{1}_{I_i}(x) \ , \tag{3.1}$$

where, for $i = 0, 1, \ldots, n$,

$$I_i = \begin{cases} (-\infty, t_1), & \text{if } i = 0 \\ [t_n, \infty), & \text{if } i = n \\ [t_i, t_{i+1}), & \text{otherwise} \end{cases} \ ,$$

and $\mu_i \in \mathbb{R}$ and $\theta_i \in \mathbb{R}$ satisfy:

1. *If $i > 0$, then $\mu_i = \sum_{j=1}^{i} t_j(\theta_{j-1} - \theta_j)$ (continuity constraints).*
2. *$\theta_0 \geq \theta_1 \geq \ldots \geq \theta_n \geq 1$ (curvature & monotonicity constraints).*

Finally, let $\tilde{\phi} \in \mathbf{\Phi}$ and take $\psi : (0,1) \to \mathbb{R}$ given by $\psi(x) = -\log \tilde{\phi}(x)$. Then

$$\phi(x) = \exp\left[-p\left(\psi(x)\right)\right] \tag{3.2}$$

is an Archimedean generator.

Proof. One can check $\phi(1) = \lim_{x \to 1^-} \phi(x) = 0$ and ϕ is strictly decreasing: ψ is strictly increasing and $x \mapsto -p(x)$ strictly decreasing (p is strictly increasing, because it is continuous and the slopes θ_i's are positive).

To check the convexity requirement, it helps to express ϕ as

$$\phi(x) = q(\tilde{\phi}(x)) \ , \tag{3.3}$$

where $q : (0, \infty) \to (0, \infty)$ is given by

$$q(x) = \sum_{i=0}^{n} e^{-\mu_i} x^{\theta_i} \mathbb{1}_{e^{-I_i}}(x) \ , \tag{3.4}$$

and

$$e^{-I_i} = \{e^{-x} : x \in I_i\} = \begin{cases} (e^{-t_1}, \infty), & \text{if } i = 0 \\ (0, e^{-t_n}], & \text{if } i = n \\ (e^{-t_{i+1}}, e^{-t_i}], & \text{otherwise} \end{cases} \ .$$

Clearly, q is convex, since the exponents $\theta_n \leq \theta_{n-1} \leq \ldots \leq \theta_0$ increase as x does and they are all greater than or equal to 1. Then, ϕ is the composition of two convex functions and q is monotonically increasing, so ϕ is convex. □

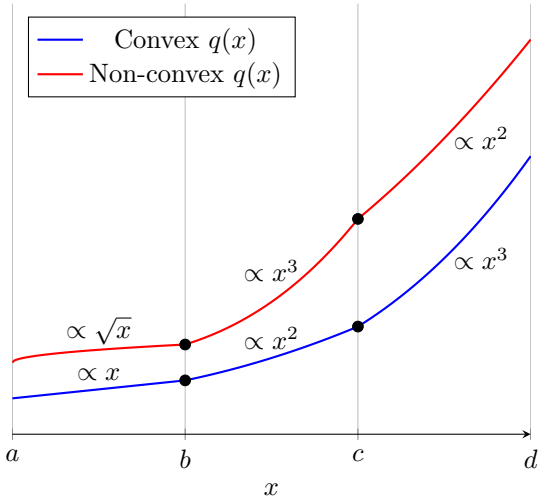


Figure 3.1: Convex and non-convex instances of the $q(x)$ function in (3.3). The convex example is proportional at each piece to x^θ, for $\theta \geq 1$ increasing as x does. The non-convex example is not convex for two reasons. On the one hand, $q(x) - K \propto \sqrt{x}$ in $[a, b]$, concave, for a certain $K \geq 0$. On the other hand, the exponent θ in the power x^θ does not increase as in the convex example: $\theta = 3$ in $[b, c]$, whereas $\theta = 2$ in $[c, d]$.

Note that the actual value of μ_0 is irrelevant. Different values produce p functions (3.1) that differ by a constant and, therefore, determine the same Archimedean copula. On the other hand, equation (3.3) shows that ϕ is strict, i.e., $\phi(0) = \infty$, if and only if $\tilde{\phi}$ is strict.

The construction in Proposition 3.1.2 is an extension or generalization of that of Proposition 3.1.1, introducing not just one, but potentially many parameters θ_i's, as Remark 3.1.1 points out below. By optimizing these constrained parameters, we can fit such a generator to actual data, providing a legitimate (Archimedean) copula estimation method. Figure 3.1 illustrates the requirement that $\theta_0 \geq \theta_1 \geq \ldots \geq \theta_n \geq 1$ in the context of the simpler (3.3) representation.

Remark 3.1.1. In Proposition 3.1.2, if $\theta_i = \theta$ for $i = 0, 1, \ldots, n$ and some $\theta \geq 1$, then $\phi(x) = (\tilde{\phi}(x))^\theta$. In particular, if $\tilde{\phi}(x) = -\log x$, the independence copula generator, then ϕ produces the Gumbel copula with parameter θ.

According to Vandenhende and Lambert, *"That copula is indeed a semi-parametric (or highly parameterized) Archimedean copula, having parameters that quantify the dependence locally"* . We believe the term *semiparametric* does not exactly correspond to *highly parameterized*, as this citation suggests. For the Archimedean copula estimation problem, *semiparametric* usually means an attempt to approximate a suitable functional parameter: the ϕ generator or an equivalent form of this. If the modelling strategy is successful enough, a few parameters might suffice.

Perhaps the foremost drawback of Vandenhende and Lambert's construction is the fact that the resulting ϕ generator is not differentiable. The smoothness

requirement is not arbitrary at all, even for modelling purposes, as the following theorem demonstrates Nelsen (2006).

Theorem 3.1.1. *Let $\phi \in \mathbf{\Phi}$ and (U, V) be distributed as C_ϕ. For $\alpha \in (0, 1)$, define the α-level curve of C_ϕ as*

$$L_\alpha(C_\phi) = \{(u, v) \in [0, 1]^2 : C_\phi(u, v) = \alpha\} \tag{3.5}$$
$$= \{(u, v) \in [0, 1]^2 : \phi(u) + \phi(v) = \phi(\alpha)\} \ . \tag{3.6}$$

Then

$$\mathbb{P}_{C_\phi}\left(L_\alpha(C_\phi)\right) = \mathbb{P}\left((U, V) \in L_\alpha(C_\phi)\right) = \phi(\alpha)\left(\frac{1}{\phi'(\alpha^-)} - \frac{1}{\phi'(\alpha^+)}\right) \ .$$

According to Theorem 3.1.1, any non-differentiable ϕ generator produces a singular[2] two-dimensional distribution, which, consequently, has no density. This poses some issues, preventing us from using maximum likelihood estimation; standard r.v. simulation techniques (Embrechts et al., 2003) still work, though[3]. Next, we present our very first example of non-trivial (with distinct θ_i's) Vandenhende and Lambert's generator.

Example 3.1.1. Consider the piecewise generator

$$\phi(x) = \begin{cases} (-\log x)^2, & \text{if } 0 < x \leq 1/e \\ -\log x, & \text{if } 1/e < x < 1 \end{cases} \ ,$$

which can be expressed as $\phi(x) = \exp\left(-p\left(-\log\left(\tilde{\phi}(x)\right)\right)\right)$, where

$$p(x) = \begin{cases} 2x, & \text{if } x \leq 0 \\ x, & \text{if } x > 0 \end{cases} \ ,$$

and $\tilde{\phi}(x) = -\log x$. Hence, ϕ complies with the conditions in Proposition 3.1.2 and is a legitimate generator. In Figure 3.2 we can visualize a simulation of the copula generated by ϕ. As we can see, there is a singular component at $L_{1/e}(C_\phi) = \{(\lambda, 1/e\lambda) : \lambda \in [1/e, 1]\}$, with $\mathbb{P}_{C_\phi}\left(L_{1/e}(C_\phi)\right) = 1/2e \approx 0.1839$.

Figure 3.3 shows the plot of a piecewise generator similar to the one presented in Example 3.1.1. For this particular generator, there would be two singular components over $L_\alpha(C_\phi)$, for $\alpha = 1/3$ and $\alpha = 2/3$; non-differentiability at these points is visually perceptible.

The practical importance of singular components is out of discussion. Many copula models, notably Marshall-Olkin copulas (Embrechts et al., 2003), incorporate them. However, the question here is whether they are a suitable resource for a general data fitting problem, rather than a problem prerequisite; after all, singular components may be detected by simply inspecting data. We believe that, for the vast majority of applications, we cannot reasonably expect singular components to apply.

[2] $L_\alpha(C_\phi)$ in (3.5) is a curve, thus it has Lebesgue measure (in $[0, 1]^2$) zero. However, it accumulates a non-null C_ϕ-probability.

[3] Partial derivatives of C_ϕ exist almost everywhere (see Remark 2.1.1).

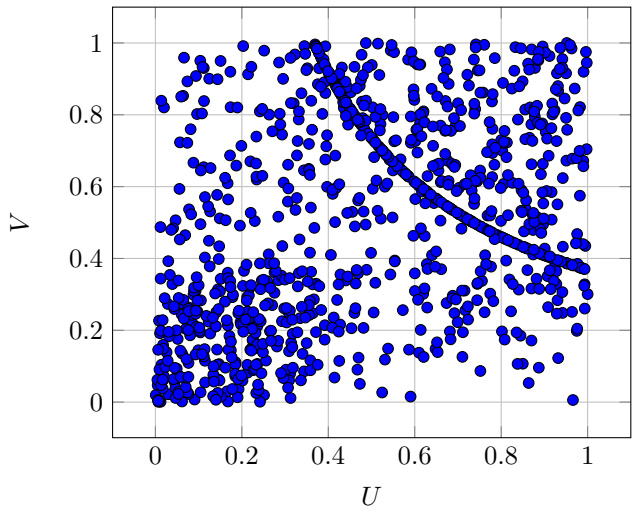


Figure 3.2: A random simulation of the copula in Example 3.1.1. Up to 191 out of 1000 sample instances (U_i, V_i) satisfy $U_i V_i \approx {}^1/_e$, where $\approx$ stands for Julia's *isapprox* routine, and, thus, lie on the singular component. The *isapprox* routine roughly checks that half of the significand digits are equal, which is a reasonable approximation to make up for the machine rounding errors.

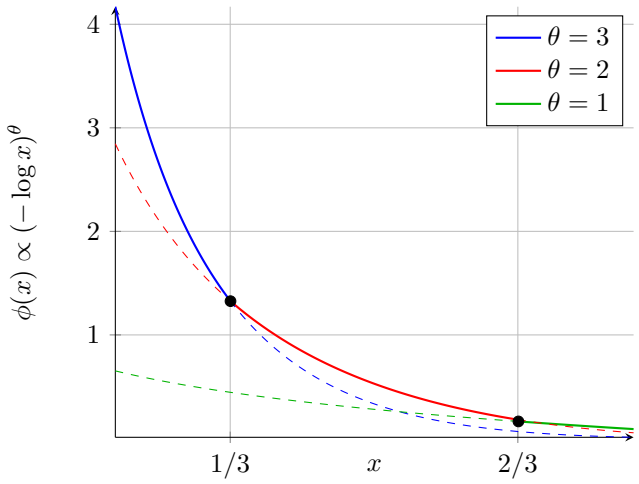


Figure 3.3: Example of Vandenhende and Lambert's piecewise generator ϕ with seed $\tilde{\phi}(x) = -\log x$. The plot shows x's in the range $[0.2, 0.8]$. The resulting generator (solid lines) comprises three pieces, joining at $x = 1/3$ and $x = 2/3$, and coloured according to the exponent θ applied. Dotted lines indicate the continuation of each piece, so that every isolated colour represents a Gumbel generator with parameter θ (proportional to $(-\log x)^\theta$).

The resulting fitted model is more rigid than it might seem. If none of the seed generator $\tilde{\phi}$ powers[4] provides a good fit, parameters will surely be distinct, thus yielding a non-differentiable ϕ which, in turn, introduces singular components; arguably a paradox, for singular components were not supposed to appear, in light of a previous data inspection. Later on, we will discuss the smoothing strategy proposed by Vandenhende and Lambert, which partially minimizes this effect.

Relation to dependence measures

The meaning of Vandenhende and Lambert's construction can be appreciated through the dependence measures introduced in Chapter 2. As we shall see, the seed generator $\tilde{\phi}$ is, to a great extent, responsible for the following properties.

Parameters θ_i naturally arise when calculating the Kendall's tau for Vandenhende and Lambert's Archimedean copula.

Proposition 3.1.3. *Let $\tilde{\phi} \in \mathbf{\Phi}$ be such that $\tau(C_{\tilde{\phi}}) = 0$. Let ϕ as in Proposition* 3.1.2*, with parameters $\theta_0, \theta_1, \ldots, \theta_n$. We have*

$$\tau(C_\phi) = \sum_{i=0}^{n} w_i \frac{\theta_i - 1}{\theta_i} \ , \tag{3.7}$$

with weights satisfying $w_i > 0$ and $\sum_{i=0}^{n} w_i = 1$.

Proof. Note that

$$0 = \tau(C_{\tilde{\phi}}) = 1 + 4 \int_0^1 \frac{\tilde{\phi}(x)}{\tilde{\phi}'(x^+)} \, dx = 1 - \sum_{i=0}^{n} w_i \ ,$$

where the weights

$$w_i = -4 \int_{\psi^{-1}(I_i)} \frac{\tilde{\phi}(x)}{\tilde{\phi}'(x^+)} \, dx \tag{3.8}$$

are positive[5]. Now, differentiating from the right (3.2),

$$\frac{\phi(x)}{\phi'(x^+)} = -\frac{1}{p'(\psi(x)^+)} \cdot \frac{1}{\psi'(x^+)} = \frac{1}{p'(\psi(x)^+)} \cdot \frac{\tilde{\phi}(x)}{\tilde{\phi}'(x^+)} \ , \tag{3.9}$$

since p' is piecewise constant in each I_i, namely $p'(x) = \theta_i$ if $x \in I_i$,

$$\begin{aligned}
\tau(C_\phi) &= 1 + 4 \int_0^1 \frac{\phi(x)}{\phi'(x^+)} \, dx \\
&= 1 + 4 \sum_{i=0}^{n} \int_{\psi^{-1}(I_i)} \frac{\phi(x)}{\phi'(x^+)} \, dx \\
&= 1 + 4 \sum_{i=0}^{n} \frac{1}{\theta_i} \int_{\psi^{-1}(I_i)} \frac{\tilde{\phi}(x)}{\tilde{\phi}'(x^+)} \, dx \\
&= 1 - \sum_{i=0}^{n} \frac{w_i}{\theta_i}
\end{aligned}$$

and the result follows from $\sum_{i=0}^{n} w_i = 1$. □

[4] See Remark 3.1.1.

[5] The integrand in (3.8) is negative, because $\tilde{\phi}(x) > 0$ and $\tilde{\phi}'(x^+) < 0$, for all $x \in (0, 1)$.

Vandenhende and Lambert present a particular case of Proposition 3.1.3 for $\tilde{\phi}(x) = -\log x$, the independence copula generator. However, this is not the only Archimedean copula with Kendall's tau equal to zero; for instance, say $\tilde{\phi}(x) = (1-x)^2$. The following remark provides an insightful interpretation of (3.7).

Remark 3.1.2. In Proposition 3.1.3, if $\theta_i = \theta$ for $i = 0, 1, \ldots, n$ and some $\theta \geq 1$, then

$$\tau(C_{\tilde{\phi}^\theta}) = \frac{\theta - 1}{\theta} \ , \tag{3.10}$$

where $\tilde{\phi}^\theta(x) = (\tilde{\phi}(x))^\theta$. Using (3.10), in a more general setting with several distinct θ_i's, we can express (3.7) as

$$\tau(C_\phi) = \sum_{i=0}^{n} w_i \tau(C_{\tilde{\phi}^{\theta_i}}) \ . \tag{3.11}$$

Using the connection between regular variation and tail dependence in Archimedean copulas (Proposition 2.4.6), the latter can be easily verified.

Proposition 3.1.4. *Let $\tilde{\phi} \in \mathbf{\Phi}$ be the seed generator for ϕ, as in Proposition* 3.1.2. *Then:*

1. *If $\tilde{\phi} \in \mathcal{R}_{0^+,\alpha}$, for $-\infty < \alpha \leq 0$, then $\phi \in \mathcal{R}_{0^+,\beta}$, with $\beta = \theta_0 \alpha$.*
2. *If $\tilde{\phi} \in \mathcal{R}_{1^-,\alpha}$, for $1 \leq \alpha < \infty$, then $\phi \in \mathcal{R}_{1^-,\beta}$, with $\beta = \theta_n \alpha$.*

Proof. It easily follows from the alternative representation (3.3):

$$\lim_{x \to 0^+} \frac{\phi(\lambda x)}{\phi(x)} = \left(\lim_{x \to 0^+} \frac{\tilde{\phi}(\lambda x)}{\tilde{\phi}(x)} \right)^{\theta_0} \ ,$$

and

$$\lim_{x \to 0^+} \frac{\phi(1-\lambda x)}{\phi(1-x)} = \left(\lim_{x \to 0^+} \frac{\tilde{\phi}(1-\lambda x)}{\tilde{\phi}(1-x)} \right)^{\theta_n} \ .$$

□

Some important consequences derive from Proposition 3.1.4.

Remark 3.1.3. Let α and β be the indices of regular variation at 0^+ of $\tilde{\phi}$ and ϕ, respectively, as in Proposition 3.1.4. Since $\theta_0 \geq 1$, we always have $\beta \leq \alpha$. Moreover, if $\alpha = 0$, then $\beta = 0$, irrespective of the value of the θ_0 parameter.

Remark 3.1.4. Let α and β be the indices of regular variation at 1^- of $\tilde{\phi}$ and ϕ, respectively, as in Proposition 3.1.4. Since $\theta_n \geq 1$, we always have $\beta \geq \alpha$.

To put things into perspective, the last two remarks reveal some limitations of Vandenhende and Lambert's semiparametric approach as regards tail dependence, according to Proposition 2.4.6. Depending on the actual indices or regular variation of the seed generator, some tail dependence indices become unreachable for C_ϕ, the Archimedean copula generated by ϕ. For instance, if $\tilde{\phi}$ is regularly varying at 0^+, with index $\alpha < 0$, then Vandenhende and Lambert's technique does not allow C_ϕ to have a lower tail dependence index in $[0, 2^{1/\alpha})$.

Similarly, if $\tilde{\phi}$ is regularly varying at 1^- with index $\alpha > 1$, then C_ϕ cannot have an upper tail dependence index in $[0, 2 - 2^{1/\alpha})$.

For this reason, the selection of the seed generator $\tilde{\phi}$ is of capital importance. Vandenhende and Lambert propose taking $\tilde{\phi}(x) = -\log x$, the independence copula generator, thus $\psi(x) = -\log(-\log x)$, on the grounds of Remark 3.1.1 and the fact that the Kendall's tau of the Gumbel family members can be interpreted as in (3.11). However, this choice prevents any attempt to capture the lower tail index.

Estimation and smoothing procedures

In spite of the fact that we cannot rely on a density to apply maximum likelihood estimation to the parameters in Proposition 3.1.2, Vandenhende and Lambert envision an alternative procedure based on an empirical estimator of the Kendall distribution (Genest and Rivest, 1993).

Let $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$ be a sample from a bivariate distribution F, not necessarily having uniform margins. To estimate the copula of the aforementioned distribution, Vandenhende and Lambert propose to minimize the *distance* between the empirical estimator of the Kendall distribution, $\hat{K}_\mathcal{D}$, and the Kendall distribution of the generator (3.2). Remember that the former is given by

$$\hat{K}_\mathcal{D}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[V_i, \infty)}(x) \ ,$$

with the empirical estimate of $F(X_i, Y_i)$

$$V_i = \frac{\#\{(X_j, Y_j) \in \mathcal{D} : X_j < X_i, \ Y_j < Y_i\}}{N - 1} \ ,$$

whereas the latter can be easily derived from (3.9):

$$K_{\boldsymbol{\beta}}(x) = x + \frac{1}{\psi'(x^+)} \sum_{i=0}^{n} \beta_i \mathbb{1}_{I_i}(\psi(x)) \ , \tag{3.12}$$

where the i-th element of vector $\boldsymbol{\beta}$ is $\beta_i = 1/\theta_i$. Obviously, optimizing the parameter vector $\boldsymbol{\beta}$ is equivalent to doing so with $\boldsymbol{\theta}$. Using all the above, the optimization problem is stated as

$$\arg\min_{\boldsymbol{\beta} \in \Theta} \left\{ \sum_{i=1}^{N} \left(\hat{K}_\mathcal{D}(V_i) - K_{\boldsymbol{\beta}}(V_i) \right)^2 \right\} \ , \tag{3.13}$$

over the parameter space

$$\Theta = \left\{ \boldsymbol{\beta} \in \mathbb{R}^{n+1} : 0 < \beta_0 \leq \beta_1 \leq \ldots \leq \beta_n \leq 1 \right\} \ . \tag{3.14}$$

Problem (3.13) can be reformulated as

$$\arg\min_{\boldsymbol{\beta} \in \Theta} \| \boldsymbol{\kappa} - \Psi \boldsymbol{\beta} \|_2^2 \ , \tag{3.15}$$

where $\boldsymbol{\kappa} \in \mathbb{R}^N$ has i-th component $\boldsymbol{\kappa}_i = \hat{K}_\mathcal{D}(V_i) - V_i$ and $\Psi \in \mathcal{M}_{N \times (n+1)}(\mathbb{R})$ has (i, j)-entry, for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n+1$,

$$\Psi_{ij} = \frac{\mathbb{1}_{I_{j-1}}(\psi(V_i))}{\psi'(V_i^+)} = \begin{cases} 1/\psi'(V_i^+), & \text{if } \psi(V_i) \in I_{j-1} \\ 0, & \text{otherwise} \end{cases} \ .$$

Now, problem (3.15) can be approached using quadratic programming, solving

$$\underset{A\boldsymbol{\beta}\geq\mathbf{b}}{\arg\min}\left\{\frac{1}{2}\boldsymbol{\beta}^T Q\boldsymbol{\beta}+\mathbf{c}^T\boldsymbol{\beta}\right\}\ , \tag{3.16}$$

where:

- $Q=\Psi^T\Psi\in\mathcal{M}_{(n+1)\times(n+1)}(\mathbb{R})$.
- $\mathbf{c}=-\Psi^T\boldsymbol{\kappa}\in\mathbb{R}^{n+1}$.
- The constraint $A\boldsymbol{\beta}\geq\mathbf{b}$ must be component-wisely interpreted, for $A\in\mathcal{M}_{(n+1)\times(n+1)}(\mathbb{R})$ mapping $\boldsymbol{\beta}$ to $(\beta_1-\beta_0,\beta_2-\beta_1,\ldots,\beta_n-\beta_{n-1},-\beta_n)^T$ and $\mathbf{b}=(0,0,\ldots,0,-1)^T\in\mathbb{R}^{n+1}$.

Observe that $A\boldsymbol{\beta}\geq\mathbf{b}$ does not impose the positivity constraint β needs to belong to Θ (3.14). This is actually the original proposal in (Vandenhende and Lambert, 2005). It is easy to add one more non-strict inequality constraint requiring $\beta_0\geq\epsilon$, for some $\epsilon\gtrsim 0$. This would always guarantee that we get a valid Archimedean generator, while maintaining a high degree of generality, but Vandenhende and Lambert do not attach any importance to this detail. We can only guess that, if the $\boldsymbol{\beta}$ we obtain by solving (3.16) does not belong to Θ, then our data set $\mathcal{D}$ is not likely to arise from a distribution whose copula is Archimedean. Looking at (3.12), since $\psi'(x^+)>0$, if any of the β_i is negative, then $K_{\boldsymbol{\beta}}(x_*^-)<x_*$ for some $x_*\in(0,1)$, which means ϕ is not an Archimedean generator, for Theorem 2.4.2 (Genest and Rivest, 1993).

Vandenhende and Lambert propose to use a large number of parameters corresponding to equidistant knots (for instance, they set $n=100$ in an example with real data) and then make up for the flexibility excess by forcing proximate parameters β_i not to be *very* different from each other. This can be achieved by adding a penalty term to the quadratic programming problem (3.16). To formalize such a proximity smoothness notion, we introduce the following definition.

Definition 3.1.1 (**r-th order differences).** Let $\mathbf{x}\in\mathbb{R}^m$. We define the r-th order difference linear map $\Delta_m^r:\mathbb{R}^m\to\mathbb{R}^{m-r}$ as

$$\Delta_m^r\mathbf{x}=\begin{cases}\Delta_{m-r+1}(\Delta_m^{r-1}\mathbf{x}), & \text{if } r\geq 2\\ \Delta_m\mathbf{x}, & \text{if } r=1\end{cases}\ ,$$

where the first order difference $\Delta_m:\mathbb{R}^m\to\mathbb{R}^{m-1}$ is given by

$$\Delta_m\mathbf{x}=(x_2-x_1,x_3-x_2,\ldots,x_m-x_{m-1})^T$$

if $\mathbf{x}=(x_1,x_2,\ldots,x_m)^T$.

From now on, we shall refer to both the linear map and its matrix with respect to the canonical basis of $\mathbb{R}^m$ and $\mathbb{R}^{m-r}$ indistinctly as Δ_m^r.

Back to the smoothing question, the formal way to force such a proximity smoothness is by targeting a low value of the squared 2-norm $\|\Delta_{n+1}^r\boldsymbol{\beta}\|^2$. For example, Vandenhende and Lambert choose a second order ($r=2$) difference

smoothing. The version of (3.16) with an added penalty $\lambda \geq 0$ to the difference norm looks like

$$\underset{A\boldsymbol{\beta}\geq\mathbf{b}}{\arg\min}\left\{\frac{1}{2}\boldsymbol{\beta}^T(Q+\lambda P)\boldsymbol{\beta}+\mathbf{c}^T\boldsymbol{\beta}\right\}\ ,$$

where $P = (\Delta^r_{n+1})^T\Delta^r_{n+1}$ The larger the penalty factor λ is, the more weight is attached to smoothing; the smaller λ is, the closer the result will remain to the original data (and its noise). The optimal λ is chosen by means of a cross-validation-type procedure.

3.1.2 A cubic spline generator model

In this section we look at a newer proposal (Lambert, 2014) that extends that of (Vandenhende and Lambert, 2005), which was reviewed in the previous section, and has 'superior properties' as compared with (Lambert, 2007). The new Archimedean copula approximation approach follows the same overall philosophy as its predecessor: not modelling the ϕ generator directly, but rather finding a g function that is *equivalent* to ϕ and easier to approximate.

Just like in Proposition 3.1.2, Lambert proposes a one-to-one correspondence

$$\phi(x) = \exp\left[-g\left(S(x)\right)\right] \tag{3.17}$$

between ϕ and a new generator $g : \mathbb{R} \to \mathbb{R}$, which plays the role of its piecewise linear counterpart p in (3.2), through a link function $S : (0, 1) \to \mathbb{R}$, which, in turn, plays the role of ψ. On this occasion, Lambert gives no choice but to take $S(x) = -\log(-\log x)$, an instance of the ψ link function with seed $\tilde{\phi}(x) = -\log x$, the independence copula generator.

The key and characteristic step of Lambert's method is the approximation of the g generator. In (Vandenhende and Lambert, 2005), g's linear counterpart, p, had to comply with several inequality constraints involving the slope of each linear piece, that is, the derivative p', Basically, p' is a piecewise flat function, not necessarily continuous, that is required to be greater or equal than 1 at each point and monotonically decreasing. From this point of view, it comes as no surprise that (Lambert, 2014) focuses on modelling g'.

Namely, Lambert proposes to build g' as a linear combination of cubic B-splines[6] associated with equidistant knots on $[S(\epsilon), S(1-\epsilon)]$, where ϵ represents a small quantity like 10^{-6}. More precisely, the $(k+1)$-th component of the equidistant knots vector of length $N+1$ is

$$t_k = S(\epsilon) + k\frac{S(1-\epsilon) - S(\epsilon)}{N}\ , \tag{3.18}$$

for $k = 0, 1, \ldots, N$. The use of smooth cubic B-splines implies g' is continuous and, consequently, ϕ is differentiable. This means the resulting copula will have a density, in contrast to (Vandenhende and Lambert, 2005).

The coordinates or coefficients that multiply the basis functions are required to be greater than or equal to 1, so that $g'(s) \geq 1$, for all s.[7] On the other

[6] There are several versions of B-spline functions, depending on the smoothness required. Here, we shall suppose Lambert refers to twice continuously differentiable B-splines.

[7] This is a sufficient condition, but not necessary. Let I and J be a partition of the indices set $\{1, 2, \ldots, n\}$. One can easily find examples of coordinates $\theta_1, \theta_2, \ldots, \theta_n$, with $\theta_i < 1$ and $\theta_j \geq 1$, for $i \in I$ and $j \in J$, such that $\sum_{k=1}^n \theta_k B_k(x) \geq 1$ for all x that belongs to the domain of the B-splines. See (de Boor and Daniel, 1974).

hand, Lambert does not explicitly specify how the resulting cubic spline function should continue out of $[S(\epsilon), S(1-\epsilon)]$, simply stating

$$g'(s) = \sum_{k=1}^{K} (1+\theta_k^2) B_k(s) \ , \tag{3.19}$$

for K B-splines $B_1, B_2, \ldots, B_K$ and corresponding parameters $\theta_1, \theta_2, \ldots, \theta_K \in \mathbb{R}$. In this context, Lambert does mention that, if $\theta_i = \theta$, for $i = 1, 2, \ldots, K$ and some $\theta \in \mathbb{R}$, then $g'(s) = 1 + \theta^2$ for all $s \in \mathbb{R}$, which would imply $\phi(x) = (-\log x)^{\zeta}$, the Gumbel copula generator with parameter $\zeta = 1 + \theta^2$.

Containing the Gumbel copula family as a particular case actually reduces the number of feasible interpretations of (3.19). To begin with, such a statement ensures that any feasible continuation must depend on some of the parameters θ_k. Perhaps the most sensible strategy is to extend (3.19) with its Taylor polynomials at $S(\epsilon)$ and $S(1-\epsilon)$. However, we soon realize that any non-flat polynomial extension, depending on the actual values of the parameters, might even produce $g'(s) < 0$, so that ϕ would not be strictly decreasing[8].

All in all, the following is a consistent extension of (3.19) to $\mathbb{R}$:

$$g'(s) = \begin{cases} \sum_{k=1}^{K} (1+\theta_k^2) B_k(s), & \text{if } S(\epsilon) \leq s \leq S(1-\epsilon) \\ 1 + \theta_1^2, & \text{if } s < S(\epsilon) \\ 1 + \theta_K^2, & \text{if } s > S(1-\epsilon) \end{cases} \ . \tag{3.20}$$

Note that if $\theta_i = \theta$, for $i = 1, 2, \ldots, K$ and some $\theta \in \mathbb{R}$, then $g'(s) = 1 + \theta^2$ for all $s \in \mathbb{R}$, using that $\sum_{k=1}^{K} B_k(s) = 1$, for all $s \in [S(\epsilon), S(1-\epsilon)]$, according to Proposition A.2.1. Also, from the piecewise definition (3.20) it follows that the resulting Archimedean generator in (3.17) is always strict, i.e., $\phi(0) = \infty$.

Lambert's claim that (3.19), plugged into (3.17), provides a valid Archimedean generator for every $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K) \in \mathbb{R}^K$ is not completely accurate. Notwithstanding, it is remarkable the fact that, under Lambert's assumptions, the parameter space for which (3.19) is a valid Archimedean generator can be made arbitrarily *large*. We shall next present a counterexample that suggests that this is in fact what Lambert actually means.

First of all, let us write down the first and second derivatives of S:

$$S'(x) = \frac{-1}{x \log x} \ , \ S''(x) = \frac{1 + \log x}{(x \log x)^2} \ .$$

Note that, for all $x \in (0, 1)$,

$$\phi'(x) = -\underbrace{\phi(x)}_{>0} \cdot \underbrace{g'(S(x))}_{\geq 1} \cdot \underbrace{S'(x)}_{>0} < 0 \ , \tag{3.21}$$

hence ϕ is strictly decreasing. However, a brief look at ϕ'' reveals we may have $\phi''(x) < 0$ if there are no additional constraints, meaning ϕ might be neither convex, nor consequently a legitimate Archimedean generator. To simplify calculations, let us introduce the λ generator (Lambert, 2007, 2014)

$$\lambda(x) = \frac{\phi(x)}{\phi'(x)} = \frac{-1}{g'(S(x)) S'(x)} \ . \tag{3.22}$$

[8] See (3.21) below.

With this new generator, as pointed out in (Lambert, 2007),

$$\phi''(x) = \frac{\phi'(x)}{\lambda(x)}(1 - \lambda'(x)) \ , \tag{3.23}$$

and thus $\phi''(x) \geq 0$ if and only $\lambda'(x) \leq 1$. Strictly speaking, λ is not differentiable at every $x \in (0,1)$, in general, as a result of the piecewise definition (3.20). Notwithstanding, λ' is piecewise continuous and one-sided derivatives of λ do exist, so that, even if $\lambda'(x_0)$ does not, having $\lambda'(x_0^+) > 1$ implies $\lambda'(x^*) > 1$, for some $x^* > x_0$.

Differentiating (3.22), we get

$$\lambda'(x) = \frac{g''(S(x))S'(x) + g'(S(x))S''(x)}{\left[g'(S(x))S'(x)\right]^2} \ . \tag{3.24}$$

The following example supposes the B-splines of g' are not defined to the left of 0, which is not Lambert's original plan, but serves as an illustrative example.

Example 3.1.2. Suppose the two left-most knots of the B-splines in (3.20) are 0 and $3/e$. Using Proposition *A*.2.6, we get $g'(0) = 1 + \theta_1^2$ and

$$g''(0^+) = 3 \cdot \frac{(1+\theta_1^2) - (1+\theta_2^2)}{0 - \frac{3}{e}} = e(\theta_2^2 - \theta_1^2) \ .$$

On the other hand, plugging $S(1/e) = 0$, $S'(1/e) = e$ and $S''(1/e) = 0$ into (3.24),

$$\lambda'\left(1/e^+\right) = \frac{1}{e}\frac{g''(0^+)}{(g'(0))^2} = \frac{\theta_2^2 - \theta_1^2}{(1+\theta_1^2)^2} \ .$$

Taking, for instance, $\theta_1 = 0$ and $\theta_2 = \sqrt{2}$, we get $\lambda'\left(1/e^+\right) = 2 > 1$. Therefore, for (3.23), the resulting ϕ generator cannot be convex and, hence, does not produce a copula.

The next counterexample follows Lambert's original setting proposal and uses the same principle as Example 3.1.2.

Example 3.1.3. Consider the $(N+1)$-knot vector (3.18) and take $\delta = (t_1 - t_0)/3$. Using Proposition *A*.2.6 and plugging $g'(S(\epsilon)) = 1 + \theta_1^2$ and

$$g''(S(\epsilon)) = 3 \cdot \frac{(1+\theta_1^2) - (1+\theta_2^2)}{t_0 - t_1} = \frac{\theta_2^2 - \theta_1^2}{\delta}$$

into (3.24), we get

$$\lambda'(\epsilon^+) = \frac{(\theta_2^2 - \theta_1^2)S'(\epsilon) + \delta(1+\theta_1^2)S''(\epsilon)}{\delta\left[(1+\theta_1^2)S'(\epsilon)\right]^2} \ .$$

Taking θ_2 sufficiently large, we secure $\lambda'(\epsilon^+) > 1$. For a more explicit counterexample, let us first take $\theta_1 = 0$. Then, since $S''(\epsilon) = (1 + \log\epsilon)(S'(\epsilon))^2$,

$$\lambda'(\epsilon^+) = \frac{\theta_2^2}{\delta S'(\epsilon)} + 1 + \log\epsilon \ .$$

Finally, realizing $-S'(\epsilon)\log\epsilon = 1/\epsilon$, we have $\lambda'(\epsilon^+) > 1$ if and only if

$$|\theta_2| > \sqrt{\frac{\delta}{\epsilon}} = \sqrt{\frac{S(1-\epsilon) - S(\epsilon)}{3\epsilon N}} \ . \tag{3.25}$$

Lambert suggests $\epsilon = 10^{-6}$ and employing $K = 11$ B-splines. For Proposition $A.1.1$ applied to twice differentiable cubic splines, we have $K = n + 2$, where $n = N + 1$ is the number of knots, thus $N = 8$. Substituting this values, we roughly have that, for $|\theta_2| > 828$, ϕ is not convex and consequently does not produce a copula.

As Example 3.1.3 and (3.25) suggest, the validity of Lambert's statement grows as $\epsilon \to 0^+$. Normally, if the technique produces an invalid generator, the latter will be so by a narrow margin and it will suffice to take a smaller ϵ to solve the issue. Beyond a reasonable threshold $\epsilon \gtrsim 0$, if the optimization process keeps producing some extreme values as parameters, then we may assume that this technique is not well-suited to our dataset.

In Section 3.1.2 we will talk about the smoothing strategy used in (Lambert, 2007, 2014), which also relies on the r-th order differences of the parameter vector (see Section 3.1.1), just like in (Vandenhende and Lambert, 2005). This certainly makes it really unlikely to find cases like Example 3.1.3 in practice.

Relation to dependence measures

The construction by Lambert and its generator g' are closely related to Kendall's tau. This connection can explicitly be stated, using Proposition 2.4.5, as

$$\begin{aligned}
\tau(C_\phi) &= 1 + 4\int_0^1 \lambda(x)\ dx \\
&= 1 - 4\int_0^1 \frac{1}{g'(S(x))S'(x)}\ dx \\
&= 1 - 4\int_{-\infty}^{\infty} \frac{\gamma(s)}{g'(s)}\ ds \ ,
\end{aligned}$$

where $\gamma(s) = \exp\left[-2(s + \exp(-s))\right]$. Note that, in contrast to (Vandenhende and Lambert, 2005), we lose track of how much of the total $\tau(C_\phi)$ each parameter θ_i (or equivalently, its transformed version $1 + \theta_i^2$) accounts for (see Proposition 3.1.3). This is not necessarily a bad feature. As a matter of fact, we can expect individual parameters to play a less prominent role than that of the overall generator as techniques transition to less parametric approaches.

The use of the link function $S(x) = -\log(-\log x)$ strongly influences the tail dependence behaviour of the Archimedean generator, just like the selection of $\tilde{\phi}(x) = -\log x$ as seed generator did in Proposition 3.1.4.

Proposition 3.1.5. *Let ϕ be as in (3.17), in one-to-one correspondence with a g generator whose derivative is expressed as (3.20). We have $\phi \in \mathcal{R}_{0^+,0} \cap \mathcal{R}_{1^-,\beta}$, for $\beta = 1 + \theta_K^2$.*

Proof. Regular variation at 0^+ does not depend on a particular polynomial continuation of g'. Since ϕ is strictly decreasing, we will use the equivalent

characterization of regular variation given in Proposition $B.0.8$. Taking $g'(s) = a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0$, for $s < S(\epsilon)$, we have

$$\begin{aligned}
\lim_{x\to 0^+} \frac{x\phi'(x)}{\phi(x)} &= \lim_{x\to 0^+} \frac{g'(S(x))}{\log x} \\
&= \sum_{k=0}^{n} a_k \left[\lim_{x\to 0^+} \frac{(-\log(-\log x))^k}{\log x} \right] \\
&= \sum_{k=0}^{n} a_k (-1)^{k+1} \left[\lim_{x\to \infty} \frac{(\log x)^k}{x} \right] \\
&= 0 \ .
\end{aligned}$$

Using the flat extension of g' to the right of $S(1-\epsilon)$ given in (3.20),

$$\begin{aligned}
\lim_{x\to 1^-} \frac{(1-x)\phi'(x)}{\phi(x)} &= \lim_{x\to 1^-} \frac{(1-x)g'(S(x))}{x\log x} \\
&= \lim_{x\to 1^-} g'(S(x)) \cdot \lim_{x\to 1^-} \frac{-1}{1+\log x} \\
&= -(1+\theta_K^2) \ .
\end{aligned}$$

□

Proposition 3.1.5 says that Lambert's technique does not allow to model lower tail dependence, since all Archimedean copulas produced are independent in the lower tail. On the contrary, this technique has no limitation as regards upper tail dependence: no upper tail index is out of reach.

Figure 3.4 illustrates Lambert's technique limitations to model a well-known Archimedean family with lower tail dependence: the Clayton family. The generator g'_θ for the Clayton copula with parameter θ,

$$g'_\theta(s) = \frac{\theta \exp(-s)}{1 - \exp(-\theta \exp(-s))} \ ,$$

cannot be approximated with polynomials as $s \to -\infty$.

Estimation and smoothing procedures

(Lambert, 2014) provides two general procedures for the estimation of the parameters θ_i in 3.20. Both rely on the density of the Archimedean copula, which, despite g' not being differentiable at $S(\epsilon)$ and $S(1-\epsilon)$, is defined almost everywhere[9] on $[0,1]^2$.

The first methods is, of course, maximum likelihood estimation. Given a sample $\mathcal{D} = \{(U_i, V_i)\}_{i=1}^N$, we want to find the $\boldsymbol{\theta} \in \mathbb{R}^K$ parameter vector which maximizes the log-likelihood function

$$\log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) = \log \left(\prod_{i=1}^{N} c_{\phi_{\boldsymbol{\theta}}}(U_i, V_i) \right) = \sum_{i=1}^{N} \log c_{\phi_{\boldsymbol{\theta}}}(U_i, V_i) \ ,$$

[9] Considering that real numbers in a computer cannot be expressed with infinite precision, we may ignore this detail.

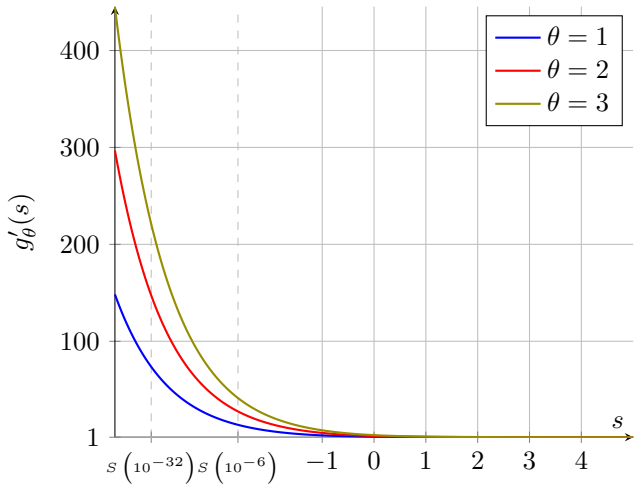


Figure 3.4: Lambert's generator $g'_\theta(s) = \frac{\theta \exp(-s)}{1-\exp(-\theta \exp(-s))}$ for the Clayton family, where $S(10^{-6}) \approx -2.6257$ and $S(10^{-32}) \approx -4.2997$.

where $c_{\phi_{\boldsymbol{\theta}}}$ represents the density of the Archimedean copula generated by $\phi_{\boldsymbol{\theta}}$ in (3.17), parametrized by $\boldsymbol{\theta}$. The model derived from (3.20) is non-identifiable, since inverting the sign of any θ_i makes no difference, but this poses no practical issue at all.

Lambert, just like in (Vandenhende and Lambert, 2005), chooses a large number of parameters, like $K = 11$, and then takes some measures to prevent overfitting, penalizing the squared 2-norm of the r-th order differences of $\boldsymbol{\theta}$, that is,

$$\|\Delta_K^r \boldsymbol{\theta}\|^2 = \boldsymbol{\theta}^T P \boldsymbol{\theta} \ , \tag{3.26}$$

where $P = (\Delta_K^r)^T \Delta_K^r$ and Δ_K^r is as in Definition 3.1.1. In (Lambert, 2014), for instance, a third order penalty, $r = 3$, is used. In the end, the objective function is a penalized log-likelihood, i.e., the target is

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^K}{\arg\max} \left\{ \log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) - \lambda \boldsymbol{\theta}^T P \boldsymbol{\theta} \right\} \ , \tag{3.27}$$

where the penalty $\lambda \geq 0$ is selected using cross-validation.

We can translate (3.27) into Bayesian terms by introducing a prior density $p(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$. The penalty λ in (3.27) has no room in a Bayesian setting, but can be sensitively incorporated in the design of $p(\boldsymbol{\theta})$ as a conditioning r.v. The derivation of such a prior can be traced back to (Lambert, 2007) and corresponds to a version of a prior for $\Delta_K^r \boldsymbol{\theta}$, which is derived from the following two hypotheses:

- The r-th order difference vector $\Delta_K^r \boldsymbol{\theta}$, of length $d = K - r$, given penalty λ, follows a multivariate normal distribution with independent components, zero mean and variances that are inversely proportional to λ. More precisely,

$$\Delta_K^r \boldsymbol{\theta} | \lambda \sim \mathcal{N}_d \left(\boldsymbol{\mu} = \mathbf{0}_d, \Sigma = \frac{1}{\lambda} I_d \right) \ ,$$

where I_d is the identity matrix of order d and $\mathbf{0}_d$ is the d-dimensional null vector. This leads to the conditional density

$$p(\Delta_K^r \boldsymbol{\theta}|\lambda) = (2\pi)^{-d/2} \cdot \lambda^{d/2} \cdot \exp\left(-\frac{\lambda}{2}\|\Delta_K^r \boldsymbol{\theta}\|^2\right) . \tag{3.28}$$

- Penalty λ follows a Gamma distribution with shape hyper-parameter a and rate hyper-parameter b, $\lambda \sim \mathcal{G}(a, b)$, with density

$$p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} , \tag{3.29}$$

where Γ is the gamma function and $a > 0$, $b > 0$. Therefore, we have $\mathbb{E}(\lambda) = a/b$.

The idea is to fix sensible values for a and b at this stage and to perform *no* subsequent optimization on them. Of course, multiple choices can be made, but some are better than others. For instance, a fairly conservative selection is $a = 1$, which implies that the mode of (3.29) is $\lambda = 0$. Taking $a \ll 1$ would be even more conservative, since $p(\lambda)$ rapidly diverges as $\lambda \to 0^+$ and thus a small right-neighbourhood of $\lambda = 0$ would account for most of the probability. By contrast, the hyper-parameter b does not seem as meaningful as a when modelling prior knowledge, apart from changing the location of the distribution. Both (Lambert, 2007, 2014) use the simplifying assumption that $a = b$.

From the previous steps we can obtain $p(\Delta_K^r \boldsymbol{\theta})$ by integrating the product of the densities (3.28) and (3.29) with respect to λ:

$$\begin{aligned} p(\Delta_K^r \boldsymbol{\theta}) &= \int_0^\infty p(\Delta_K^r \boldsymbol{\theta}|\lambda) \cdot p(\lambda)\ d\lambda \\ &= (2\pi)^{-d/2} \frac{b^a}{\Gamma(a)} \cdot \int_0^\infty \lambda^{a+\frac{d}{2}-1} \cdot \exp\left[-\lambda\left(b + \frac{1}{2}\|\Delta_K^r \boldsymbol{\theta}\|^2\right)\right]\ d\lambda . \\ &= (2\pi)^{-d/2} \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma\left(a + \frac{d}{2}\right)}{\left(b + \frac{1}{2}\|\Delta_K^r \boldsymbol{\theta}\|^2\right)^{a+d/2}} \end{aligned} \tag{3.30}$$

The last derivation reveals that the r-th order difference vector follows a multivariate t-distribution, namely

$$\Delta_K^r \boldsymbol{\theta} \sim t_d\left(\nu = 2a, \boldsymbol{\mu} = \mathbf{0}_d, \Sigma = \frac{b}{a} I_d\right) . \tag{3.31}$$

The form of the matrix Σ in (3.31) supports the assumption $a = b$, resulting in a simple identity matrix. Note also that, for a multivariate t-distribution, $\nu \leq 2$ (equivalently $a \leq 1$) implies covariances are undefined.

Once removed the effect of the penalty λ, we can see the components of $\Delta_K^r \boldsymbol{\theta}$ are not independent any more, just uncorrelated. Another interesting remark relates to the posterior distribution of λ, which can be calculated using Bayes'

theorem:

$$\begin{aligned} p(\lambda|\Delta_K^r\boldsymbol{\theta}) &= \frac{p(\Delta_K^r\boldsymbol{\theta}|\lambda)\cdot p(\lambda)}{p(\Delta_K^r\boldsymbol{\theta})} \\ &= \frac{\left(b+\frac{1}{2}\|\Delta_K^r\boldsymbol{\theta}\|^2\right)^{a+d/2}}{\Gamma\left(a+\frac{d}{2}\right)}\cdot\lambda^{a+\frac{d}{2}-1}\cdot\exp\left[-\lambda\left(b+\frac{1}{2}\|\Delta_K^r\boldsymbol{\theta}\|^2\right)\right] , \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \end{aligned} \tag{3.32}$$

where

$$\begin{aligned} \alpha &= a+\frac{d}{2} \\ \beta &= b+\frac{1}{2}\|\Delta_K^r\boldsymbol{\theta}\|^2 \end{aligned} . \tag{3.33}$$

Observe that both the prior (3.29) and the posterior (3.32) distributions belong to the Gamma family and how the knowledge about the r-th order differences updates the penalty λ and its original hyper-parameters, a and b, to the α and β in (3.33). The latter is best appreciated when looking at the conditional expectation of (3.32), that is

$$\mathbb{E}\left(\lambda|\Delta_K^r\boldsymbol{\theta}\right) = \frac{\alpha}{\beta} = \frac{a+\frac{d}{2}}{b+\frac{1}{2}\|\Delta_K^r\boldsymbol{\theta}\|^2} ,$$

from which we can draw two conclusions:

- Fixed dimensionality d, the larger the overfitting $\|\Delta_K^r\boldsymbol{\theta}\|^2$, the smaller we can expect the penalty to be and vice versa.
- Fixed the overfitting $\|\Delta_K^r\boldsymbol{\theta}\|^2$, the larger the dimensionality d, the larger we can expect the penalty to be and vice versa.

All the previous remarks suggest that the prior (3.30) is sensible. Hence, Lambert translates it to $\boldsymbol{\theta}\in\mathbb{R}^K$ as

$$p(\boldsymbol{\theta}) \propto \frac{1}{\left(b+\frac{1}{2}\|\Delta_K^r\boldsymbol{\theta}\|^2\right)^{a+d/2}} = \frac{1}{\left(b+\frac{1}{2}\boldsymbol{\theta}^T P\boldsymbol{\theta}\right)^{a+d/2}} ,$$

where $P=(\Delta_K^r)^T\Delta_K^r$. All in all, the Bayesian counterpart of (3.27) is the maximum *a posteriori* estimator

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^K}{\arg\max}\left\{\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})\cdot p(\boldsymbol{\theta})\right\} = \underset{\boldsymbol{\theta}\in\mathbb{R}^K}{\arg\max}\left\{\frac{\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})}{\left(b+\frac{1}{2}\boldsymbol{\theta}^T P\boldsymbol{\theta}\right)^{a+d/2}}\right\} ,$$

or equivalently, since log is monotonically increasing and $\operatorname{rank}(P) = \operatorname{rank}(\Delta_K^r) = K-r=d$,

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^K}{\arg\max}\left\{\log\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) - \left(a+\frac{\operatorname{rank}(P)}{2}\right)\log\left(b+\frac{1}{2}\boldsymbol{\theta}^T P\boldsymbol{\theta}\right)\right\} , \tag{3.34}$$

that is the mode of the posterior distribution of $\boldsymbol{\theta}$ given the sample $\mathcal{D}$. It turns out to be interesting to compare (3.34) with the non-Bayesian approach (3.27). As we can see, the regularization term in (3.34) has a built-in dimensionality factor that involves rank(P), whereas in (3.27) one has to *manually* tune the penalty λ to suit each particular parameter setting. Note that this term is consistent with the remark that the penalty should go up as the dimensionality of the problem does, given a fixed *overfitting* value.

In this review, we have made use of Bayesian statistics to find a reasoned and reasonable regularization term for the original objective function, the log-likelihood. If the reader is interested in posterior distribution sampling using Markov chain Monte Carlo, we refer them to (Lambert, 2014).

Notes and comments

In this section, we have reviewed two Archimedean copula modelling strategies. Both embrace a semiparametric philosophy, with different degrees of development, and the use of polynomial splines. While (Vandenhende and Lambert, 2005) sees semiparametric approaches as highly parametrized, its proposal turned out to be less flexible than expected. The proposal in (Lambert, 2014), though, succeeds in accomplishing the flexibility goal with fewer parameters and effectively removing the singular components issue that harmed (Vandenhende and Lambert, 2005). As regards dependence measures, neither (Vandenhende and Lambert, 2005) nor (Lambert, 2014) offer a full and satisfactory solution. In particular, tail dependence is not addressed in any of these papers.

It is important to realize the huge step from (Vandenhende and Lambert, 2005) to (Lambert, 2014) and, at the same time, praise the conciseness and efficiency of the former. In the next section, we will look at the extension of the cubic spline technique in (Lambert, 2014) to conditional copulas.

3.2 Conditional copula modelling

Given a sample $\mathcal{D} = \left\{(U_1^i, U_2^i, U_3^i)\right\}_{i=1}^N$ from a 3-dimensional r.v. (U_1, U_2, U_3) with uniform margins and distributed according to a certain copula C, Definition 2.6.1 says we can attempt to re-construct C in a two-phase estimation process:

1. Firstly, estimate two of the bivariate margin copulas, for instance: $C_{1,2}$, linking U_1 and U_2, and $C_{2,3}$, linking U_2 and U_3. Let us call C_{12} and C_{23} their respective estimates, arising from any suitable bivariate copula estimation technique. For instance, any of the flexible semiparametric techniques introduced in the previous section.

2. Secondly, for each $u_2 \in [0,1]$, estimate the copula $C_{1,3|2}(*,*|u_2)$ corresponding to the 2-dimensional r.v. $[(U_1, U_3)|U_2 = u_2]$, where the r.v. U_2 is called the *covariate*. The latter conditional r.v. has margins $[U_1|U_2 = u_2]$ and $[U_3|U_2 = u_2]$, which are the conditional margins U_1 given $U_2 = u_2$ and U_3 given $U_2 = u_2$, respectively. These two margins have the following CDFs:

$$\begin{matrix} [U_1|U_2 = u_2] \sim C_{1|2}(*|u_2) \\ [U_3|U_2 = u_2] \sim C_{3|2}(*|u_2) \end{matrix},$$

where

$$\begin{aligned} C_{1|2}(*|u_2) &= \partial_2 C_{1,2}(*, u_2) = \mathbb{P}(U_1 \leq *|U_2 = u_2) \\ C_{3|2}(*|u_2) &= \partial_1 C_{3,2}(*, u_2) = \mathbb{P}(U_3 \leq *|U_2 = u_2) \end{aligned} .$$

Hence, the CDF of $[(U_1, U_3)|U_2 = u_2]$, which is of course $(u_1, u_3) \mapsto \partial_2 C(u_1, u_2, u_3)$, can be obtained using Sklar's theorem:

$$\partial_2 C(u_1, u_2, u_3) = C_{1,3|2}(C_{1|2}(u_1|u_2), C_{3|2}(u_3|u_2)|u_2) \ .$$

Therefore, the original copula C can be expressed as

$$\begin{aligned} C(u_1, u_2, u_3) &= \int_0^{u_2} \partial_2 C(u_1, \upsilon, u_3) \ d\upsilon \\ &= \int_0^{u_2} C_{1,3|2}(C_{1|2}(u_1|\upsilon), C_{3|2}(u_3|\upsilon)|\upsilon) \ d\upsilon \end{aligned} .$$

If, additionally, C is absolutely continuous with density c, then

$$\begin{aligned} c(u_1, u_2, u_3) = c_{1,3|2}(C_{1|2}(u_1|u_2), C_{3|2}(u_3|u_2)|u_2) \cdot \\ \cdot c_{1,2}(u_1, u_2) \cdot c_{2,3}(u_2, u_3) \end{aligned} , \tag{3.35}$$

where $c_{1,3|2}(*, *|u_2)$ is the density of the conditional copula $C_{1,3|2}(*, *|u_2)$ and $c_{1,2}$ and $c_{2,3}$ are the densities of the bivariate margins $C_{1,2}$ and $C_{2,3}$, respectively. If we suppose $C_{1,2}$ and $C_{2,3}$ to be well approximated by the estimates C_{12} and C_{23} from the previous step, we can get rid of the bivariate margin factors $c_{1,2}(u_1, u_2) \cdot c_{2,3}(u_2, u_3)$ in (3.35) and focus on estimating, for each $u_2 \in [0, 1]$, the conditional copula $C_{1,3|2}(*, *|u_2)$.

As pointed out in (Acar et al., 2012), we can always try to estimate C by maximizing directly the log-likelihood (3.35), but this is computationally more demanding. If we hypothesize a model for the conditional copula family $\{C^{\boldsymbol{\alpha}}_{13|2}(*, *|u_2)\}_{u_2 \in [0,1]}$ depending on the vector of parameters $\boldsymbol{\alpha} \in \mathbb{R}^{n_1}$ and two models for the bivariate margins $C^{\boldsymbol{\beta}}_{12}$ and $C^{\boldsymbol{\gamma}}_{23}$, for $\boldsymbol{\beta} \in \mathbb{R}^{n_2}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{n_3}$, then the resulting model for C is $C_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} = (\boldsymbol{\alpha}; \boldsymbol{\beta}; \boldsymbol{\gamma}) \in \mathbb{R}^n$ and the total number of parameters is $n = n_1 + n_2 + n_3$. The density of $C_{\boldsymbol{\theta}}$ is

$$\begin{aligned} c_{\boldsymbol{\theta}}(u_1, u_2, u_3) = c^{\boldsymbol{\alpha}}_{13|2}\left(\partial_2 C^{\boldsymbol{\beta}}_{12}(u_1, u_2), \partial_1 C^{\boldsymbol{\gamma}}_{23}(u_2, u_3)|u_2\right) \cdot \\ \cdot c^{\boldsymbol{\beta}}_{12}(u_1, u_2) \cdot c^{\boldsymbol{\gamma}}_{23}(u_2, u_3) \end{aligned} . \tag{3.36}$$

Therefore, the log-likelihood function to maximize, given sample $\mathcal{D}$, is

$$\log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) = \sum_{i=1}^{N} \log c_{\boldsymbol{\theta}}(U_1^i, U_2^i, U_3^i) \ . \tag{3.37}$$

Since the bottleneck of the estimation process turns out to be the computation of the gradient of (3.37) with respect to $\boldsymbol{\theta}$, which has to be performed at each iteration of the optimization algorithm if complex semiparametric models are used, we may be unable to afford computing the n partial derivatives, for a large n. Also, note that these gradient computations are not light either, because

those parameters belonging to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ appear as arguments of $c^{\boldsymbol{\alpha}}_{13|2}$, rising computational costs.

The two-phase vine estimation process gives us the opportunity to apply complex semiparametric models for both the conditional copula family and the bivariate margins, using a large number of parameters and smoothing procedures in both cases. Assuming estimates C_{12} and C_{23}, with densities c_{12} and c_{23} respectively, we can replace (3.36) with

$$c_{\boldsymbol{\theta}}(u_1, u_2, u_3) = c^{\boldsymbol{\theta}}_{13|2}\left(\partial_2 C_{12}(u_1, u_2), \partial_1 C_{23}(u_2, u_3)|u_2\right) \cdot \\ \cdot c_{12}(u_1, u_2) \cdot c_{23}(u_2, u_3) \quad , \tag{3.38}$$

in (3.37), since optimization for the the bivariate margins is not needed any longer. In (3.38) all parameters $\boldsymbol{\theta}$ are devoted to modelling the conditional copula family. What is more, the values $\partial_2 C_{12}(U^i_1, U^i_2)$ and $\partial_1 C_{23}(U^i_2, U^i_3)$ need not be computed every time. Instead, we can maximize

$$\log \mathcal{L}^*(\boldsymbol{\theta}|\mathcal{D}) = \sum_{i=1}^{N} \log c^{\boldsymbol{\theta}}_{13|2}(U_i, V_i|W_i) \ , \tag{3.39}$$

where the datasets $\mathcal{D}_{U,V} = \{(U_i, V_i)\}_{i=1}^N$ and $\mathcal{D}_W = \{W_i\}_{i=1}^N$, for

$$\begin{aligned} U_i &= \partial_2 C_{12}(U^i_1, U^i_2) \\ V_i &= \partial_1 C_{23}(U^i_2, U^i_3) \ , \\ W_i &= U^i_2 \end{aligned}$$

can be precomputed. For the above reasons, from now on we will use the two-phase vine estimation procedure. Observe that, because the covariate is uniformly distributed, each conditional copula is actually a 3-copula. Hence, (3.39) is a legitimate log-likelihood function.

In this section we will look at several strategies to model the conditional copula family $\{C_{1,3|2}(*, *|u_2)\}_{u_2 \in [0,1]}$. The first two techniques we will introduce, (Acar et al., 2012) and (Lopez-Paz et al., 2013), make use of known parametric copula families to build their conditional copulas and focus on the estimation of the effect of the covariate u_2 on the family member selection. This effect is modelled in a non-parametric way as a function (Acar et al., 2012) and as a stochastic process (Lopez-Paz et al., 2013). By contrast, the third and last technique, (Lambert, 2014), *builds* the copula family itself in a semiparametric manner, extending the bivariate Archimedean copula modelling strategy discussed in the previous section. To the best of our knowledge, (Lambert, 2014) has been the only approach of this kind so far. Our proposal, which will be thoroughly discussed in Chapter 4, relies on a similar extension of a bivariate copula modelling method, namely (Hernández-Lobato and Suárez, 2011).

Fully non-parametric approaches like (Gijbels et al., 2011) make hardly any assumption on the underlying conditional copula. Roughly speaking, the copula estimate does not fit data instances, but rather data instances build the copula itself. Though interesting, we shall not address them in this work.

3.2.1 A first order local polynomial model of the covariate

The approach in (Acar et al., 2011) assumes all conditional copulas belong to the same parametric family. Let $\{c_\theta\}_{\theta \in \Theta}$ be a parametric family of copula

densities, with parameter space $\Theta \subset \mathbb{R}$. One could estimate a suitable parameter function $\theta : [0,1] \to \Theta$ mapping covariate values to density parameters. To ensure θ ranges in Θ, it is commonplace to model an alternative latent function $\eta : [0,1] \to \mathbb{R}$ and then build θ as

$$\theta(w) = g^{-1}(\eta(w)) \ , \tag{3.40}$$

where $g^{-1} : \mathbb{R} \to \Theta$ is known as the inverse link function.

As pointed out in (Acar et al., 2011), it is possible to model η parametrically, for instance, considering n-th degree polynomials. Following (3.39), given the calibration function

$$\eta_{\boldsymbol{\beta}}(w) = \beta_0 + \beta_1 w + \beta_2 w^2 + \ldots + \beta_n w^n \ , \tag{3.41}$$

parametrized by the coefficients vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_n)$, we can always target

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^{n+1}}{\arg\max} \left\{ \sum_{i=1}^{N} \log c \left[U_i, V_i; g^{-1}(\eta_{\boldsymbol{\beta}}(W_i)) \right] \right\} \ ,$$

where we have written $c_\theta(u,v)$ as $c(u,v;\theta)$ to enhance readability. According to Acar et al., the underlying calibration function η might not be well approximated by the polynomial $\eta_{\boldsymbol{\beta}}$ or any other predefined parametric functional form. Even though we agree with this statement, it must be stressed that a semiparametric approach using polynomial splines, which is not discussed in (Acar et al., 2011), might provide a better fit. Nevertheless, both parametric and semiparametric approaches to modelling the covariate effect are relatively simple when considered in the context of parametric conditional copulas. For this reason, we shall not further discuss them in this work.

Estimation procedure

The approach used in (Acar et al., 2011) is actually non-parametric. Given a covariate value w, our goal is to select a suitable parameter θ, regardless of the way we represent η in (3.40). We can even give up on the idea of a global definition of η. (Acar et al., 2011) hypothesize that η is smooth and thus can be locally approximated with polynomials, even if we cannot attain a complete definition of η. This way, given a covariate instance $W_i \in \mathcal{D}_W$, we can approximate η at W_i with its n-th degree Taylor polynomial around w, that is

$$\eta(W_i) \approx \sum_{k=0}^{n} \frac{\eta^{(k)}(w)}{k!} (W_i - w)^k = \eta_{\boldsymbol{\beta}}(W_i - w) \ , \tag{3.42}$$

where $\eta_{\boldsymbol{\beta}}$ is the polynomial calibration function in (3.41) with coefficients $\beta_k = \eta^{(k)}(w)/k!$, for $k = 0, 1, \ldots, n$. If we could find a suitable $\eta_{\boldsymbol{\beta}}$, that is, if the approximation (3.42) were *sufficiently* accurate for those $W_i \in \mathcal{D}_W$ that are close *enough* to w, then we would have a good estimate for $\theta(w)$, which is of course $g^{-1}(\beta_0)$. Note that coefficients other than $\beta_0 \approx \eta(w)$ are not needed any more. Now, we still have to specify a goodness of fit measure for $\boldsymbol{\beta}$ and decide which instances in $\mathcal{D}_W$ are eligible for the estimation. These aspects are considered in the local likelihood estimation problem

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^{n+1}}{\arg\max} \left\{ \sum_{i=1}^{N} K_h(W_i - w) \cdot \log c \left[U_i, V_i; g^{-1}(\eta_{\boldsymbol{\beta}}(W_i - w)) \right] \right\} \ , \tag{3.43}$$

where $K_h(x) = hK(x/h)$ and $K : \mathbb{R} \to [0, \infty)$ is a kernel function like

$$K(x) = \frac{3}{4(1-x^2)} \mathbb{1}_{(-1,1)}(x) \ ,$$

the Epanechnikov kernel. Roughly speaking, (3.43) is an average of conditional copula log-densities in which the closer W_i is to w, the more weight is attached to the i-th term. The rationale behind this design lies in the fact that the approximation (3.42) is more reliable when W_i is closer to w. The closeness notion is conveyed through the bandwidth parameter h.

The number of data instances with a perceptible contribution to the local likelihood estimation might be really small. The scarcity of data prevents us from using a high degree n for the local polynomial approximation, due to the *curse of dimensionality*. Therefore, (Acar et al., 2011) employs a first order approximation:

$$\underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{\arg\max} \left\{ \sum_{i=1}^{N} K_h(W_i - w) \cdot \log c \left[U_i, V_i; g^{-1}(\beta_0 + \beta_1(W_i - w)) \right] \right\} \ . \quad (3.44)$$

Optimal bandwidth selection

It is difficult to determine beforehand a good value for the bandwidth h: it depends on the dataset $\mathcal{D}$ itself. Acar et al. propose a leave-one-out cross-validation to tune h. Let $\mathcal{D} = \{(U_i, V_i, W_i)\}_{i=1}^{N}$ be the available data and let $\theta_h(w|\mathcal{D})$ be the parameter estimate for the covariate w arising from (3.44), considering the sample D and the bandwidth h. Now, define $\theta_h^{(-i)}(w) = \theta_h\left(w|\mathcal{D}^{(-i)}\right)$, where $\mathcal{D}^{(-i)} = \mathcal{D} \setminus \{(U_i, V_i, W_i)\}$. An *optimal* bandwidth h^* can be obtained through

$$h^* = \underset{h\in\mathcal{H}}{\arg\max} \left\{ \sum_{i=1}^{N} \log c \left[U_i, V_i; g^{-1}\left(\theta_h^{(-i)}(W_i)\right) \right] \right\} \ ,$$

where the bandwidth search space $\mathcal{H} \subset (0, 1]$ admits both a discrete and a continuous specification, depending on the optimization algorithm one wants to use.

Strengths and weaknesses

Being non-parametric, one disadvantage of the approach in (Acar et al., 2011) is the lack of an actual training phase or model construction phase. As mentioned earlier, non-parametric methods do not build models, but rather use available data as a model itself. For each new covariate w we have to perform the optimization (3.44). This may be unaffordable for real-time applications.

Another drawback, which is specific to (Acar et al., 2011), is the strong limitation of using parametric copulas. The approach we will review next, (Lopez-Paz et al., 2013), partially overcomes this issue by using mixtures of parametric copulas.

On the positive side, (Acar et al., 2011) assumes as few hypotheses as possible on the calibration function, which leads to low-bias solutions. That makes this approach a valuable asset when proposing and selecting alternative models.

For instance, it may provide us with a *sketch* of the covariate dependence structure, which could eventually help determining: tuning parameters for other methods, initialization values, appropriate basis functions for semiparametric representations and so on. In Lambert's words:

> "Nonparametric versions are desirable to suggest or to validate parametric specifications, or even as a substitute for these models."
>
> (Lambert, 2014)

3.2.2 A Gaussian process model of the covariate

The approach in (Lopez-Paz et al., 2013), though non-parametric too, differs from (Acar et al., 2011) in the way they estimate the effect of the covariate. Actually, their proposal can theoretically deal with an arbitrarily large number of covariates. We shall present here a particular instantiation of their method for the 3-vine case.

The construction

They first consider a parametric family of copula densities $\{c_\tau\}_{\tau\in[-1,1]}$ that can be indexed in terms of any value of Kendall's τ (household parametric families like Frank's, Glayton's or Gumbel's satisfy this). Then, they build the target conditional copula density $c(*,*|w)$ as a mixture of instances of the aforementioned parametric family, where the weight of each instance is modelled as a density that takes into account the covariate w. Letting the function $\tau : \mathbb{R} \to [-1.1]$

$$\tau(x) = 2\Phi(x) - 1 \ ,$$

where $\Phi : \mathbb{R} \to (0,1)$ is the CDF of the univariate standard normal distribution, the construction can be formally stated as

$$c(u,v|w) = \int_{\mathbb{R}} c_{\tau(x)}(u,v) \cdot p(x|w)\ dx \ . \tag{3.45}$$

As in (Acar et al., 2011), there is actually no training stage: given w, the copula density $\tilde{c}(*,*|w)$ is built directly from data.

Therefore, the gist of Lopez-Paz et al.'s approach is the proposal for the density $p(x|w)$. Given a dataset $\mathcal{D} = (D_{U,V}, D_W)$, they take

$$p(x|w) = p_{X_w}(x|\mathcal{D}) \ , \tag{3.46}$$

i.e., the posterior density of a r.v. X_w from an *a priori* Gaussian process $\mathbb{X} = \{X_w\}_{w\in[0,1]}$ that is indexed by the covariate w. Hence, given the vector of covariate values $\mathbf{w} = (w_1, w_2, \ldots, w_k)$, the r.v. $\mathbb{X}_\mathbf{w} = (X_{w_1}, X_{w_2}, \ldots, X_{w_k})^T$ follows a k-variate Gaussian distribution, namely $\mathbb{X}_\mathbf{w} \sim \mathcal{N}_k(\boldsymbol{\mu}_\mathbf{w}, \Sigma_\mathbf{w})$ with mean vector $\boldsymbol{\mu}_\mathbf{w} \in \mathbb{R}^k$ and covariance matrix $\Sigma_\mathbf{w} \in \mathcal{M}_{k\times k}(\mathbb{R})$ having (i,j)-element

$$(\Sigma_\mathbf{w})_{ij} = K(w_i, w_j) = \sigma_0 + \sigma \exp\left[-\lambda(w_i - w_j)^2\right] \ ,$$

where the hyper-parameters σ_0, σ and $\lambda \geq 0$ are optimized later on in the inference process. Note that this covariance matrix expresses that two r.v. X_{w_i} and X_{w_j} are more strongly correlated the closer the covariates w_i and w_j are. On

the other hand, taking $\boldsymbol{\mu}_{\mathbf{w}} = (0, 0, \ldots, 0) \in \mathbb{R}^k$ we assume no prior knowledge on the location of the X_w's.

Now, consider the covariates vector $\mathbf{w} = (W_1, W_2, \ldots, W_N)$, whose components arise from $\mathcal{D}_W$, and the extended covariates vector $\bar{\mathbf{w}} = (w; \mathbf{w})$ formed by pre-appending an arbitrary covariate w to $\mathbf{w}$. Note that

$$\Sigma_{\bar{\mathbf{w}}} = \begin{bmatrix} k & \mathbf{k}^T \\ \mathbf{k} & \Sigma_{\mathbf{w}} \end{bmatrix} ,$$

where $k = K(w, w)$ and $\mathbf{k} = (K(w, W_1), K(w, W_2), \ldots, K(w, W_N))^T$. Then, the conditional r.v. $[X_w | \mathbb{X}_{\mathbf{w}} = \mathbf{x}]$ follows a univariate normal distribution (Cambanis et al., 1981), $\mathcal{N}(\mu, \sigma^2)$, with mean $\mu = \mu_w + (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{w}})^T \Sigma_{\mathbf{w}}^{-1} \mathbf{k} = \mu_w + \mathbf{k}^T \Sigma_{\mathbf{w}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{w}})$ and variance $\sigma^2 = k - \mathbf{k}^T \Sigma_{\mathbf{w}}^{-1} \mathbf{k}$.

Let $p_{X_w | \mathbb{X}_{\mathbf{w}}}(x|\mathbf{x})$ be the density of $[X_w | \mathbb{X}_{\mathbf{w}} = \mathbf{x}]$. Lopez-Paz et al. model (3.46) as

$$p_{X_w}(x|\mathcal{D}) = \int_{\mathbb{R}^N} p_{X_w | \mathbb{X}_{\mathbf{w}}}(x|\mathbf{x}) \cdot p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x}|\mathcal{D}) \, d\mathbf{x} . \tag{3.47}$$

Note that all inference then comes down to finding the posterior $p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x}|\mathcal{D})$, which conveys the available information $\mathcal{D}$. Of course, should we replace $p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x}|\mathcal{D})$ with $p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x})$ in (3.47), then we would simply have $p_{X_w}(x)$, the univariate normal density we *a priori* assumed, and no inference would be made. Then, all that is left is to look for a way to *update* the density of $\mathbb{X}_{\mathbf{w}}$ on the grounds of $\mathcal{D}$. To do that, Lopez-Paz et al. propose to think of each realization $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$ of $\mathbb{X}_{\mathbf{w}}$ as an input vector of parameters for the simplified conditional copula mixture model

$$c_{\mathbf{x}}(u, v|w) = \sum_{i=1}^{N} c_{\tau(x_i)}(u, v) \cdot p(x_i|w) , \tag{3.48}$$

where

$$p(x_i|W_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} ,$$

assuming the covariate w ranges only in $\mathcal{D}_W$. Hence, the covariate w acts as a density selector. If $w = W_i$, the resulting density is $c_{\tau(x_i)}$, the only one affected by the parameter x_i, which is a realization of X_{W_i}. Certainly, (3.48) *mimics* the form of the originally intended mixture model (3.45), but using some trivial 1-0 weights. The likelihood function for (3.48) has the convenient form

$$\mathcal{L}(\mathbf{x}|\mathcal{D}) = \prod_{i=1}^{N} c_{\mathbf{x}}(U_i, V_i|W_i) = \prod_{i=1}^{N} c_{\tau(x_i)}(U_i, V_i) . \tag{3.49}$$

Finally, we know the posterior of $\mathbf{x}$ given $\mathcal{D}$ is proportional to the product of the likelihood of $\mathbf{x}$ given $\mathcal{D}$ by the prior density of $\mathbf{x}$:

$$p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x}|\mathcal{D}) \propto \mathcal{L}(\mathbf{x}|\mathcal{D}) \cdot p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x}) .$$

With all that, (3.45) becomes

$$c(u, v|w) = \int_{\mathbb{R} \times \mathbb{R}^N} c_{\tau(x)}(u, v) \cdot p_{(X_w; \mathbb{X}_{\mathbf{w}})}(x; \mathbf{x}|\mathcal{D}) \, dx \, d\mathbf{x} , \tag{3.50}$$

where

$$p_{(X_w; \mathbb{X}_{\mathbf{w}})}(x; \mathbf{x}|\mathcal{D}) = p_{X_w | \mathbb{X}_{\mathbf{w}}}(x|\mathbf{x}) \cdot p_{\mathbb{X}_{\mathbf{w}}}(\mathbf{x}|\mathcal{D}) . \tag{3.51}$$

Approximation steps

In practice, we can approximate (3.45) by sampling $\{x_i\}_{i=1}^n$ from (3.46) and then averaging over copula densities $\{c_{\tau(x_i)}\}_{i=1}^n$, that is

$$c(u, v|w) \approx \frac{1}{n} \sum_{i=1}^{n} c_{\tau(x_i)}(u, v) \ . \tag{3.52}$$

Moreover, because of the way we arrived to (3.46), it might be more convenient to sample $\{(x_i; \mathbf{x}_i)\}_{i=1}^n$ from (3.51) and use the first components $\{x_i\}_{i=1}^n$ in (3.52).

As we see, evaluating (3.51) is computationally expensive because of the likelihood (3.49), specially when the number of data instances N is large. Considering that, for the approximation (3.52) to be accurate, a large number of sample points n is needed, the evaluation of $p_{(X_w;\mathbb{X}_\mathbf{w})}(x; \mathbf{x}|\mathcal{D})$ and, in particular, of $p_{\mathbb{X}_\mathbf{w}}(\mathbf{x}|\mathcal{D})$, is critical. Lopez-Paz et al. employ the Expectation Propagation (EP) algorithm to approximate $p_{\mathbb{X}_\mathbf{w}}(\mathbf{x}|\mathcal{D})$ with a computationally tractable density.

Strengths and weaknesses

The proposal in (Lopez-Paz et al., 2013) provides a very interesting and innovative way to model the effect of several covariates. When the number of covariates is rather large, Lopez-Paz et al.'s proposal is unparalleled, but the one-covariate case presented here seems comparatively simple.

As mentioned in the case of (Acar et al., 2011), being (Lopez-Paz et al., 2013) non-parametric, computational efficiency is not one of its strengths, despite the use of the EP algorithm.

A major disadvantage of (Lopez-Paz et al., 2013) is the use of parametric copulas as building blocks. Sure enough, the resulting copula, a mixture, does not belong to a parametric family in general, contrary to (Acar et al., 2011), but mixture properties depend, to a great extent, on those of the underlying copulas. Besides, even though we can always try to apply some model selection criterion, the number of known and convenient parametric families is relatively small as compared with the infinite number of semiparametric Archimedean copulas that (Lambert, 2014) and (Hernández-Lobato and Suárez, 2011) allow to model.

3.2.3 A fully semiparametric spline-based model

The methods we will look at in this section are radically different from the previous two. They were first presented in (Lambert, 2014) as an extension of the bivariate copula modelling method presented in Section 3.1.2, also included in (Lambert, 2014). Remember that Lambert, in the bivariate case, proposed to model a latent function $g' : \mathbb{R} \to \mathbb{R}$ instead of the Archimedean generator ϕ. A possible extension to the conditional case is to model $\partial_1 g : \mathbb{R} \times [0, 1] \to \mathbb{R}$, such that, for each covariate $w \in [0, 1]$,

$$\phi_w(x) = \exp\left[-g\left(S(x), w\right)\right] \tag{3.53}$$

is a legitimate Archimedean generator. This way, we would have a conditional copula family $\{C(*,*|w) = C_{\phi_w}\}_{w\in[0,1]}$, where

$$C(u,v|w) = C_{\phi_w}(u,v) = \phi_w^{-1}(\phi_w(u) + \phi_w(v)) \ .$$

A second extension approach consists in using the bivariate case technique to model an Archimedean generator ϕ that is semiparametrically and conveniently deformed under the covariate effect using a map

$$\begin{array}{rccc} T_{\boldsymbol{\theta}}: & \boldsymbol{\Phi} \times [0,1] & \longrightarrow & \boldsymbol{\Phi} \\ & (\phi, w) & \mapsto & \phi_w \end{array} \tag{3.54}$$

that is tuned according to some vector of parameters $\boldsymbol{\theta}$.

Strictly speaking, the second approach is not an extension of the bivariate case technique, but rather a more general conditional copula modelling framework where other Archimedean generators, perhaps obtained by other means, can be plugged in. For instance, a single-parameter Archimedean generator like Frank's ϕ_θ could be estimated as the core of the transformed family $\{T_{\boldsymbol{\theta}}(\phi_\theta, w)\}_{w\in[0,1]}$. However, in the latter example, the overall copula is not fully semiparametric. Also, note that the assumption of a unique underlying core generator ϕ, posed by (3.54), is more restrictive than (3.53), even if the transformation $T_{\boldsymbol{\theta}}$ is highly parametrized.

The construction

If, for each $w \in [0,1]$, $x \mapsto g(x,w)$ is modelled as described in (Lambert, 2014), then each conditional copula C_{ϕ_w} is semiparametric. Lambert's proposal to model the covariate effect is also semiparametric. We just have to think of each parameter θ_k in (3.19) as a function of the covariate w, which can also be approximated using cubic splines. Let $B_1^*, B_2^*, \ldots, B_L^*$ be the cubic B-splines corresponding to some set of knots on the covariate space $[0,1]$. We can approximate $\partial_1 g$ with

$$\frac{\partial g}{\partial s}(s,w) = \sum_{k=1}^{K} \left(1 + \theta_k(w)^2\right) B_k(s) \ , \tag{3.55}$$

where

$$\theta_k(w) = \sum_{l=1}^{L} \theta_{kl} B_l^*(w) \ , \tag{3.56}$$

for a matrix of parameters $\Theta = (\theta_{kl}) \in \mathcal{M}_{K\times L}(\mathbb{R})$.

The foremost issue of (3.56) lies in the total number of parameters: $K \times L$. Remember that Lambert recommended to set $K = 11$ in the bivariate case. Sticking to this heuristic, in the context of (3.56), would mean the total number of parameters is bounded from below by $11L$, which leaves very little flexibility for modelling the covariate. (Lambert, 2014) proposes an alternative formulation for $\partial_1 g$, the additive conditional spline Archimedean copula family, in which we replace (3.56) with

$$\theta_k(w) = \alpha_k + \sum_{l=1}^{L} \beta_l B_l^*(w) \tag{3.57}$$

in (3.55), with vectors of parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_L)$. Since $\sum_{l=1}^{L} B_l^*(w) = 1$, for Proposition A.2.1, adding a constant $\mu \in \mathbb{R}$ to every β_l is equivalent to doing so with every α_k. This causes a model identifiability problem that can be solved by adding a identification constraint like

$$\sum_{l=1}^{L} \beta_l = 0 \ . \tag{3.58}$$

Therefore, the additive model (3.57) consists of $K + L - 1$ parameters.

Regardless of the decision we make between (3.56) and (3.57), (3.55) is a smooth function both in the s and w variables. This feature partially guarantees that the conditional copula dependence varies smoothly with the covariate, much like the local linearity assumption and the Gaussian prior did in (Acar et al., 2011) and (Lopez-Paz et al., 2013), respectively.

Finally, we end this subsection introducing the alternative proposal by Lambert, based on the semiparametric transformation (3.54): the flex-power Archimedean conditional copula family. Taking ϕ_w in (3.54) defined as

$$\phi_w(x) = \left[\phi\left(x^{\alpha(w)}\right)\right]^{\beta(w)} \ , \tag{3.59}$$

for some functions $\alpha : [0,1] \to (0,1]$ and $\beta : [0,1] \to [1,\infty)$, we get a conditional family of Archimedean generators. The functions α and β are commonly known as interior and exterior power transformations, respectively.[10] Then, we complete the construction by modelling all functions ϕ, α and β in a semiparametric manner: ϕ follows the specification in Section 3.1.2 with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$, whereas

$$\alpha(w) = \left[1 + \left(\sum_{l=1}^{L} \alpha_l B_l^*(w)\right)^2\right]^{-1}$$

and

$$\beta(w) = 1 + \left(\sum_{l=1}^{L} \beta_l B_l^*(w)\right)^2 \ ,$$

for $B_1^*, B_2^*, \ldots, B_L^*$ cubic B-splines defined on $[0,1]$ and respective parameter vectors $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_L)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_L)$. The total number of parameters is $K + 2L$.

Estimation procedure and smoothing

The estimation procedure for the conditional case is analogous to the bivariate one. We can either maximize a penalized log-likelihood estimation or use Bayesian inference with a prior distribution that penalizes the overfitting. The form of the penalty differs from parametrization to parametrization, though.

For the full parametrization (3.56), the penalty targets the squared 2-norms of the r-th and s-th order differences of the rows and columns of Θ, respectively. Let $\Theta_{:,l}$ and $\Theta_{k,:}$ be the l-th column and k-th row of Θ, respectively, both

[10] Remember that the exterior power was introduced in Proposition 3.1.1, in the context of (Vandenhende and Lambert, 2005).

expressed as (column) vectors. Also, let I_K and I_L be the identity matrix of orders L and K, respectively. It can be easily verified that

$$\sum_{l=1}^{L} \|\Delta_K^r \Theta_{:,l}\|^2 = (\operatorname{vec}\Theta)^T \cdot (I_K \otimes P_K) \cdot \operatorname{vec}\Theta \ , \tag{3.60}$$

where $P_K = (\Delta_K^r)^T \Delta_K^r$ and $\otimes$ represents the Kronecker product. Similarly, transposing Θ and using (3.60),

$$\sum_{k=1}^{K} \|\Delta_L^s \Theta_{k,:}\|^2 = \left(\operatorname{vec}\Theta^T\right)^T \cdot (I_L \otimes P_L) \cdot \operatorname{vec}\Theta^T \ , \tag{3.61}$$

where $P_L = (\Delta_L^s)^T \Delta_L^s$. In order to combine (3.60) and (3.61) in a single quadratic form, we must permute the entries of $\operatorname{vec}\Theta^T$ in (3.61) to obtain $\operatorname{vec}\Theta$. Some basic results reported in (Zhang and Ding, 2013) actually allow us to write

$$\sum_{k=1}^{K} \|\Delta_L^s \Theta_{k,:}\|^2 = (\operatorname{vec}\Theta)^T \cdot (P_L \otimes I_L) \cdot \operatorname{vec}\Theta \ . \tag{3.62}$$

Finally, combining (3.60) and (3.62) with respective weights $\alpha \geq 0$ and $\beta \geq 0$, we get the penalty term

$$P_{\alpha,\beta}(\Theta) = (\operatorname{vec}\Theta)^T \left(\alpha I_K \otimes P_K + \beta P_L \otimes I_L\right) \operatorname{vec}\Theta \ . \tag{3.63}$$

which added to the log-likelihood $\log \mathcal{L}(\Theta|\mathcal{D})$ gives us the conditional copula estimation problem:

$$\underset{\Theta \in \mathcal{M}_{K\times L}(\mathbb{R})}{\arg\max} \ \left\{\log \mathcal{L}(\Theta|\mathcal{D}) - P_{\alpha,\beta}(\Theta)\right\} \ . \tag{3.64}$$

Lambert does not expand on the Bayesian counterpart of (3.64) and focuses on the Bayesian estimation of the additive parameters model. In this case, smoothing is imposed on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by penalizing their r-th and s-th order differences, respectively, using exactly the same rationale employed in Section 3.1.2. However, while the penalty matrix for $\boldsymbol{\alpha}$ is simply $P_{\boldsymbol{\alpha}} = (\Delta_K^r)^T \Delta_K^r$, the penalty matrix for $\boldsymbol{\beta}$, $P_{\boldsymbol{\beta}} = (\Delta_L^s)^T \Delta_L^s + \epsilon I_L$, includes an extra term to impose the identifiability constraint.[11] The prior densities of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are additionally assumed to be independent, giving rise to the posterior distribution

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{D}) \propto \frac{\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{D})}{\left(b_{\boldsymbol{\alpha}} + \frac{1}{2}\boldsymbol{\alpha}^T P_{\boldsymbol{\alpha}} \boldsymbol{\alpha}\right)^{a_{\boldsymbol{\alpha}} + \frac{\operatorname{rank} P_{\boldsymbol{\alpha}}}{2}} \left(b_{\boldsymbol{\beta}} + \frac{1}{2}\boldsymbol{\beta}^T P_{\boldsymbol{\beta}} \boldsymbol{\beta}\right)^{a_{\boldsymbol{\beta}} + L/2}} \ , \tag{3.65}$$

with hyper-parameters $a_{\boldsymbol{\alpha}}, b_{\boldsymbol{\alpha}}, a_{\boldsymbol{\beta}}, b_{\boldsymbol{\beta}}$.

A similar posterior can be envisioned for the flex-power family, penalizing the vector differences, too. Namely,

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{D}) \propto \frac{\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{D})}{\prod_{\boldsymbol{x} \in \{\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}\}} \left(b_{\boldsymbol{x}} + \frac{1}{2}\boldsymbol{x}^T P_{\boldsymbol{x}} \boldsymbol{x}\right)^{a_{\boldsymbol{x}} + \frac{\operatorname{rank} P_{\boldsymbol{x}}}{2}}} \ , \tag{3.66}$$

[11] This manner of imposing the identifiability constraint is theoretically questionable, in our view. As a matter of fact, we can ensure identifiability by solving for one of the β_l's in the linear constraint (3.58).

with penalty matrices $P_{\boldsymbol{x}}$ and hyper-parameters $a_{\boldsymbol{x}}$ and $b_{\boldsymbol{x}}$, for $\boldsymbol{x} \in \{\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}\}$.

The posterior densities (3.65) and (3.66) can then be used to obtain estimates for the model parameters, by finding either the posterior mean or mode. (Lambert, 2014) reports efficient estimation procedures to sample the posterior distribution.

Strengths and weaknesses

All three solutions reviewed in this section represent a big improvement with respect to (Acar et al., 2011) and (Lopez-Paz et al., 2013) as refers to flexibility. (Lambert, 2014) is the only semiparametric approach and is consequently devoid of the limitations imposed by parametric assumptions, except for the copulas being Archimedean. On the other hand, having a training phase in which to estimate parameters, computations are much more efficient than those of non-parametric methods.

These techniques inherit some of the drawbacks that the unconditional version has, though. Firstly, despite being flexible, tail dependence cannot be properly addressed. Also, note that the form of the unconditional construction does not actually suit the conditional extension. Looking at (3.55), we see that the B-splines approximation is asymmetric, due to the square power: fixed s, $w \mapsto \partial_1 g(s, w)$ is always an even-degree piecewise polynomial.

Even though we pose no question on the effectiveness of the smoothing procedures proposed by (Lambert, 2014), we still feel that the squared norm of a parameter differences vector is not the ideal semiparametric roughness measure. When smoothness is required for several vectors or a matrix of parameters, one misses having a single elegant penalty that can be expressed at an abstract function level.

A positive aspect of (Lambert, 2014) we have not expanded on in this section, since it is out of scope, is that we can easily adapt the additive conditional spline Archimedean copula family (3.57) to include additional covariates without the total number of parameters getting out of control. Remember that (Lopez-Paz et al., 2013) has this feature too, but (Acar et al., 2011) does not.

Chapter 4

Spline-based semiparametric copulas

> – Hace muchas semanas que vengo estudiando los microfilmes sobre Historia de las matemáticas –dijo Harlan–, y los libros de las distintas Realidades del Quinientos setenta y cinco. Las diferentes Realidades no tienen mucha importancia. Las matemáticas no cambian. El orden de su desarrollo tampoco cambia. No importa cómo se pueda variar una Realidad, la Historia del crecimiento de las matemáticas sigue siendo la misma. Los matemáticos han cambiado; diferentes personas han realizado los descubrimientos, pero los resultados finales son los mismos... De todas formas, he aprendido mucho. ¿Qué le parece eso?
>
> *El fin de la eternidad*
> Isaac Asimov

In this chapter we introduce a method to model Archimedean copulas, based on approximating the generator function by polynomial splines. The method is an extension of the proposal made in (Hernández-Lobato and Suárez, 2011). The novel contributions of our work are (i) to give complete proofs of some fundamental results (ii) to provide further insight into tail dependence (iii) to provide further details on estimation and implementation.

4.1 Semiparametric Archimedean copulas

The technique by Hernández-Lobato and Suárez shares the same goal and core philosophy as (Lambert, 2014). They both aim at estimating a suitable Archimedean generator from data by assuming a simple yet flexible approximate form. In both proposals the generator is approximated using splines. However, the approximations are constructed following completely different paths and it is difficult to establish how they are related.

One aspect of (Hernández-Lobato and Suárez, 2011) that sets it apart from other proposals is its special emphasis on tail dependence. As shown in Chapter

3, the approximation made in (Vandenhende and Lambert, 2005) has some limitations to capture both the lower and upper tail dependence, whereas (Lambert, 2014), though more flexible, produces copulas that are independent in the lower tail. By contrast, Hernández-Lobato and Suárez's proposal can represent any lower and upper tail dependence index.

4.1.1 The construction of the latent function

Next, we shall introduce Hernández-Lobato and Suárez's construction in several steps.

Solving the convexity constraint issue

Among the properties an Archimedean generator must comply with, arguably the most difficult to model is the convexity constraint. The following proposition-definition shows how it can be implemented at the cost of an extra integration step.

Proposition 4.1.1 (**F-generator**). *Let $F : (0,1) \to (0,\infty)$ be continuous and monotonically increasing. Then $\phi : (0,1] \to [0,\infty)$ given by*

$$\phi(x) = \int_x^1 \frac{1}{F(t)}\, dt \tag{4.1}$$

is an Archimedean generator.

Proof. ϕ is a well-defined function whose value at $x = 1$ is 0. On the other hand, ϕ is continuously differentiable with $\phi'(x) = -1/F(x) < 0$, because F is continuous. Hence, ϕ is strictly decreasing. Moreover, since F is monotonically increasing, so is ϕ' and, therefore, ϕ must be convex.[1] □

Indeed, the above construction conveniently takes care of the convexity constraint. At this stage, it is feasible to make an approximation of F based on splines that fulfils one of the remaining conditions: monotonicity. Consider the F-generator $F(x) = \sum_{i=1}^{n} \theta_i B_i(x)$, defined over $[0,1]$. According to (Schumaker, 2007), if $\epsilon \leq \theta_1 \leq \theta_2 \leq \ldots \leq \theta_n$, for some $\epsilon > 0$, then F is positive and monotonically increasing. However, this approach has two shortcomings. On the one hand, it implies that F is bounded, which is not needed in Proposition 4.1.1. On the other hand, adding all the newer constraints on the θ_i's hampers the estimation process from a computational standpoint.

The F-generator is unique except for a constant positive factor, just like ϕ: that is, if $c > 0$, then $x \mapsto cF(x)$ generates the same Archimedean copula as F.

Example 4.1.1. The Archimedean generator $\phi(x) = -\log x$, which generates the independence copula, arises from $F(x) = x$.

Note that, since F is monotonically increasing, we can extend it to both 0 and 1, even if it is not necessary that F be defined at these points to compute ϕ from (4.1). In particular, we can define $F(0) = \lim_{x\to 0^+} F(x) \geq 0$ and $F(1) = \lim_{x\to 1^-} F(x)$, where the latter limit can be infinite. We will assume this domain extension from now on and write $F(0)$ and $F(1)$ instead of $F(0^+)$ and $F(1^-)$, respectively.

[1] $\phi \in \mathcal{C}^1(0,1)$ is convex if and only if ϕ' is monotonically increasing.

Remark 4.1.1. Let F and ϕ be in the relation expressed in Proposition 4.1.1. $F(0) = 0$ if and only if $\phi'(0) = -\infty$. On the other hand, $\phi(0) = \infty$ implies $\phi'(0) = -\infty$, whereas the converse is not true.[2] Consequently, a necessary condition for ϕ to be strict is $F(0) = 0$.

Observe that F induces a *Lebesgue-Stieltjes* measure μ_F on $(0, 1)$.[3] Interestingly, the F-generator corresponding to the independence copula is $F(x) = x$, with $\mu_F = \lambda_{\mathbb{R}|[0,1]}$, that is the Lebesgue measure restricted to $[0, 1]$ or, equivalently, the probability distribution of a uniform r.v. U on $[0, 1]$ (see Example 4.1.1). An advantage of the parametrization in terms of F is that this function has an interpretation that conveys more information than ϕ. However, some ϕ's cannot be represented as in (4.1).

Proposition 4.1.2. *An Archimedean generator ϕ can be expressed in terms of an F-generator, via Proposition* 4.1.1*, if and only if $\phi \in \mathcal{C}^1(0, 1)$.*

Proof. In the proof of Proposition 4.1.1 we saw that (4.1) is continuously differentiable. Conversely, let $\phi \in \mathcal{C}^1(0, 1)$ be an Archimedean generator. Being ϕ strictly decreasing does not guarantee $\phi'(x) < 0$, $\forall x \in (0, 1)$, but the additional convexity hypothesis does, by forcing ϕ' to be monotonically increasing.[4] Therefore $F(x) = -1/\phi'(x)$ is well-defined, positive, continuous and monotonically increasing. The fundamental theorem of Calculus can then be used to derive (4.1) from these premises. □

As stated in Theorem 3.1.1, a non-differentiable Archimedean generator produces singular components. In a general data-fitting setting, singular components are rare. Therefore, we hardly lose any generality by assuming the existence of an F-generator.

Remember that regular variation is a convenient hypothesis to ensure tail-dependent behaviour. We shall now discuss how F and ϕ relate to each other under regular variation. However, to be consistent with Proposition 4.1.1, not every regularly varying F is allowed.

Proposition 4.1.3. *Let F be as in Proposition* 4.1.1 *and let $\alpha \in \mathbb{R}$. We have:*

1. *If $F \in \mathcal{R}_{0^+,\alpha}$, necessarily $\alpha \geq 0$.*
2. *If $F \in \mathcal{R}_{1^-,\alpha}$, necessarily $\alpha \leq 0$.*

Proof. The proof of both cases is based on Remark B.0.1.

1. If $\alpha < 0$, then $F(0^+) = \infty$ and F would not be monotonically increasing. Examples with $\alpha \geq 0$ are possible.
2. If $\alpha > 0$, then $F(1^-) = 0$ and F would not be monotonically increasing. Examples with $\alpha \leq 0$ are possible.

□

The following result summarizes all possible relations between the regular variation of F and ϕ.

[2] See Example 4.1.3.

[3] For all $a, b \in (0, 1)$, $a \leq b$, μ_F satisfies $\mu_F((a, b]) = F(b) - F(a)$.

[4] Then, if $\phi'(x^*) = 0$ for some $x^* > 0$, necessarily $\phi'(x) = 0$ for all $x \in (x^*, 1)$, thus ϕ would not be strictly decreasing, leading to a contradiction.

Proposition 4.1.4. *Let F and ϕ be as in Proposition* 4.1.1. *We have:*

1. *If $F \in \mathcal{R}_{0^+,\alpha}$, with $\alpha \geq 0$, then $\phi \in \mathcal{R}_{0^+,\beta}$, where $\beta = \min\{1-\alpha, 0\}$.*
2. *If $F \in \mathcal{R}_{1^-,\alpha}$, with $\alpha \leq 0$, then $\phi \in \mathcal{R}_{1^-,\beta}$, where $\beta = 1-\alpha$.*

Proof.

1. First of all, note that if $F \in \mathcal{R}_{0^+,\alpha}$, then $x \mapsto 1/F(x) \in \mathcal{R}_{0^+,-\alpha}$. Using Proposition B.0.5, there exists $x_0 > 0$ such that $1/F(x) = x^{-\alpha}\ell(x)$ for all $0 < x < x_0$.

   Suppose $\alpha > 1$. We have, for $0 < x < x_0$,

   $$\phi(x) = \phi(x_0) + \int_x^{x_0} t^{-\alpha}\ell(t)\ dt\ ,$$

   and, using Karamata's theorem (Corollary B.0.2),

   $$\int_x^{x_0} t^{-\alpha}\ell(t)\ dt \sim -\frac{x^{1-\alpha}\ell(x)}{1-\alpha}\ ,$$

   as $x \to 0^+$.

   From the above results and the characterization of regular variation for continuously differentiable functions with normalized slowly varying part (Proposition B.0.8), which holds for Archimedean generators, as we saw in (2.20), it follows that

   $$\begin{aligned}\lim_{x\to 0^+} \frac{x\phi'(x)}{\phi(x)} &= -\lim_{x\to 0^+} \frac{x}{\phi(x)F(x)} \\ &= -\lim_{x\to 0^+} \frac{x^{1-\alpha}\ell(x)}{\phi(x_0) + \int_x^{x_0} t^{-\alpha}\ell(t)\ dt} \\ &= 1-\alpha\ ,\end{aligned}$$

   where we have used that $\lim_{x\to 0^+} x^{1-\alpha}\ell(x) = \infty$, according to Remark B.0.1. Therefore, $\phi \in \mathcal{R}_{0^+,1-\alpha}$.

   Now, suppose $\alpha < 1$. Using Karamata's theorem,

   $$\phi(0) - \phi(x) = \int_0^x t^{-\alpha}\ell(t)\ dt \sim \frac{x^{1-\alpha}\ell(x)}{1-\alpha}\ ,$$

   as $x \to 0^+$. Then, from Remark B.0.1, $\lim_{x\to 0^+} x^{1-\alpha}\ell(x) = 0$, which means $\phi(0) < \infty$. Therefore, $\phi \in \mathcal{R}_{0^+,0}$.

   Finally, if $\alpha = 1$, we either have $\phi(0) < \infty$ or $\phi(0) = \infty$. In the former case, $\phi \in \mathcal{R}_{0^+,0}$, as in the case $\alpha < 1$. The same holds for the latter: using L'Hôpital's rule,

   $$\lim_{x\to 0^+} \frac{\phi(\lambda x)}{\phi(x)} = \lambda \left[\lim_{x\to 0^+} \frac{F(\lambda x)}{F(x)}\right]^{-1} = \lambda \cdot \lambda^{-1} = 1\ .$$

   To sum up, if $\alpha > 1$, then $\phi \in \mathcal{R}_{0^+,1-\alpha}$, whereas if $\alpha \leq 1$, then $\phi \in \mathcal{R}_{0^+,0}$. This can be compactly expressed as the statement we wanted to prove.

2. Regardless of the value of α, we have $\phi(1) = 0$. Therefore, using L'Hôpital's rule,

$$\lim_{x\to 0^+} \frac{\phi(1-\lambda x)}{\phi(1-x)} = \lambda \left[\lim_{x\to 0^+} \frac{F(1-\lambda x)}{F(1-x)}\right]^{-1} = \lambda \cdot \lambda^{-\alpha} = \lambda^{1-\alpha} \ .$$

□

Regular variation in F also determines, in most cases, the strictness of the ϕ generator.

Proposition 4.1.5. *Let $F \in \mathcal{R}_{0^+,\alpha}$ be an F-generator as in Proposition* 4.1.1 *and let ϕ be its corresponding Archimedean generator. We have:*

1. *If $\alpha > 1$, then $\phi(0) = \infty$.*
2. *If $\alpha < 1$, then $\phi(0) < \infty$.*
3. *If $\alpha = 1$, then $\phi(0)$ depends on the form of F.*

Proof.

1. If $\alpha > 1$, then $\phi \in \mathcal{R}_{0^+,\beta}$, where $\beta = 1 - \alpha < 0$, for Proposition 4.1.4. Then, the result follows from Remark B.0.1.
2. This relation has been derived in Proposition 4.1.4.
3. Example 4.1.1 shows a regularly varying F with index $\alpha = 1$ such that $\phi(0) = \infty$. To see an instance of the opposite case, consider $F(x) = x(\log x)^2$ as $x \to 0^+$, i.e., there exists $x_0 > 0$ such that the equality holds if $0 < x < x_0$.[5] Then,

$$\phi(0) - \phi(x_0) = \int_0^{x_0} \frac{1}{t(\log t)^2}\, dt = \left[\frac{-1}{\log t}\right]_0^{x_0} = \frac{-1}{\log x_0} < \infty \ .$$

□

Monotonicity

Assuming that F is absolutely continuous, the monotonicity constraint is fulfilled assuming the form

$$F(x) = \eta + \int_0^x f(t)\, dt \ , \tag{4.2}$$

for some $\eta \in \mathbb{R}$ and some $f : [0,1] \to \bar{\mathbb{R}}$ integrable on $[0,x]$, for all $x \in (0,1)$. It is not difficult to see that F is monotonically increasing if and only if $f(x) \geq 0$ at almost[6] every $x \in [0,1]$, which is a huge simplification.

Additionally, according to the Lebesgue differentiation theorem, F is differentiable almost everywhere, with $F'(x) = f(x)$ at almost every $x \in (0,1)$. However, not every continuous and monotonically increasing function can be expressed as (4.2).

[5] Of course, $F(x) = x(\log x)^2$ for all $x \in (0,1)$ is not monotonically increasing and thus not an F-generator.

[6] Sure enough, f must be finite at almost every point too if integrable.

Example 4.1.2 (Cantor function). The Cantor function $\mathfrak{C} : [0,1] \to [0,1]$ (Dovgoshey et al., 2006) is a renowned example of a continuous function which is not absolutely continuous. Consider the set $\mathfrak{B}([0,1])$ of all functions $F : [0,1] \to \mathbb{R}$ with $\|F\|_\infty = \sup_{x\in[0,1]} |F(x)| < \infty$. $\mathfrak{C}$ can be constructed as the limit in $(\mathfrak{B}([0,1]), \|*\|_\infty)$ of the sequence

$$F_{n+1}(x) = \begin{cases} \frac{1}{2}F_n(3x), & \text{if } 0 \le x \le \frac{1}{3} \\ \frac{1}{2}, & \text{if } \frac{1}{3} < x < \frac{2}{3} \\ \frac{1}{2} + \frac{1}{2}F_n(3x-2), & \text{if } \frac{2}{3} \le x \le 1 \end{cases} , \tag{4.3}$$

where F_1 can be arbitrarily chosen in $\mathfrak{B}([0,1])$.[7] $\mathfrak{C}$ exhibits a counter-intuitive behaviour, being monotonically increasing on the whole $[0,1]$ with $\mathfrak{C}'(x) = 0$ almost everywhere, yet having $\mathrm{Im}(\mathfrak{C}) = [0,1]$. Therefore,

$$\mathfrak{C}(1) - \mathfrak{C}(0) = 1 \neq 0 = \int_0^1 \mathfrak{C}'(x)\, dx \ .$$

Since $F(0) = \eta$, η must be non-negative. From Remark 4.1.1, $\eta = 0$ is needed to get a strict ϕ. Finally, to enforce $F(x) > 0$ for all $x > 0$, one can simply take $f(x) > 0$, for all $x \in (0,1)$. As we shall soon see, we can use this positivity constraint to our advantage when dealing with the asymptotic behaviour of f as $x \to 0^+$ and $x \to 1^-$. We could also enforce the positivity of F by expressing it in terms of B-splines[8] and requiring that the coefficients be non-negative, but practical and theoretical results (de Boor and Daniel, 1974) disapprove that approach.

Integrable functions need not be continuous in general and may even diverge, provided that they do not accumulate an infinite area around the singularity. Such is the case of

$$f(x) = \begin{cases} \frac{1}{\sqrt{|x-1/2|}}, & \text{if } x \neq \frac{1}{2} \\ \infty & \text{if } x = \frac{1}{2} \end{cases} ,$$

whose value $f(1/2) = \infty$ could actually be different, without changing the definition of F.[9]

Even though it is possible to deal with a non-continuous f in a setting like (4.2), practical reasons urge us not to do so. Estimation procedures for an Archimedean copula C_ϕ, such as *maximum likelihood estimation*, rely on the existence of a density of C_ϕ, say c_ϕ. The simplest way to ensure that the copula density is well defined is by imposing $C_\phi \in \mathcal{C}^2((0,1)\times(0,1))$, so that

$$c_\phi(u,v) = \frac{\partial^2 C_\phi}{\partial u \partial v}(u,v) = -\frac{\phi'(u)\phi'(v)\phi''(C_\phi(u,v))}{[\phi'(C_\phi(u,v))]^3} \ . \tag{4.4}$$

This is equivalent to requiring $\phi \in \mathcal{C}^2(0,1)$ and, since

$$\phi'(x) = -\frac{1}{F(x)},\ \phi''(x) = \frac{f(x)}{(F(x))^2} \ , \tag{4.5}$$

[7] $(\mathfrak{B}([0,1]), \|*\|_\infty)$ is a Banach space and the mapping $F_n \mapsto F_{n+1}$ (4.3) is a contraction (Dovgoshey et al., 2006), so the Banach fixed point theorem applies.

[8] B-splines will be our preferred modelling tool in the following sections.

[9] f could perfectly take any other value in $\mathbb{R}$ at $x = 1/2$.

also equivalent to $f \in \mathcal{C}(0,1)$.

The previous discussion and Remark 4.1.1 motivate the introduction of the following alternative latent function.

Proposition 4.1.6 (f**-generator).** *Let $f : (0,1) \to (0,\infty)$ be continuous and integrable on $[0,x]$ for all $x \in (0,1)$. Then*

$$F(x) = \int_0^x f(t)\ dt\ , \tag{4.6}$$

is an absolutely continuous F-generator, according to Proposition 4.1.1*, satisfying $F(0) = 0$.*

Again, f is unique except for a constant positive factor that comes out of integral (4.6), producing an equivalent F.

The parametrization of the Archimedean copula in terms of f, which has simpler properties than ϕ of F, involves a loss of representation capacity. Nonetheless, as shown in Corollary 4.1.1, this loss is small in the case of strict Archimedean copulas.

Proposition 4.1.7. *The Archimedean copula generator $\phi \in \mathbf{\Phi}$ can be expressed in terms of an f-generator, via Proposition* 4.1.6*, if and only if $\phi \in \mathcal{C}^2(0,1)$, $\phi'(0) = -\infty$ and $\phi''(x) > 0$ for all $x \in (0,1)$.*

Proof. It easily follows from Proposition 4.1.6 and equations (4.5) and

$$F(x) = -\frac{1}{\phi'(x)},\ f(x) = \frac{\phi''(x)}{(\phi'(x))^2}\ . \tag{4.7}$$

□

Corollary 4.1.1. *Let $\phi \in \mathbf{\Phi}_\infty$. It holds that ϕ arises from an f-generator, via Proposition* 4.1.6*, if and only if $\phi \in \mathcal{C}^2(0,1)$ and $\phi''(x) > 0$ for all $x \in (0,1)$.*

Proof. Requiring that ϕ be strict makes the condition $\phi'(0) = -\infty$ redundant in Proposition 4.1.7. □

We will now give an example of ϕ that arises from a certain f and another one in which it is not possible to express ϕ in terms of f.

Example 4.1.3. The Archimedean generator $\phi(x) = 1-\sqrt{x}$, arises from $F(x) = 2\sqrt{x}$, which, in turn, is derived from $f(x) = 1/\sqrt{x}$. Note that $F(0) = 0$ and $\phi'(0) = -\infty$, yet $\phi(0) < \infty$.

Example 4.1.4. The Archimedean generator $\phi(x) = (1-x)^\alpha$, for any $\alpha \geq 1$, does not arise from the construction in Proposition 4.1.6. Indeed, $F(x) = -1/\phi'(x) = (1-x)^{1-\alpha}/\alpha$, thus $F(0) = 1/\alpha > 0$.

The meaning of both f and F-generators we previously alluded to is stated in the following important result.

Theorem 4.1.1. *Let (U,V) be a r.v. with uniform margins on $[0,1]$ and let C_ϕ, the Archimedean copula generated by ϕ, be their joint CDF. Let f be a generator as in Proposition* 4.1.6*. Letting $W = C_\phi(U,V)$ and $Z = V$, the density of the r.v. (W,Z) is*

$$p(w,z) = \frac{f(w)}{\int_0^z f(t)\ dt}\mathbb{1}_{[0,z]}(w) = \frac{f(w)}{F(z)}\mathbb{1}_{[0,z]}(w)\ . \tag{4.8}$$

Because $Z \sim \mathcal{U}[0,1]$, the density $p(w,z)$ of (W,Z) coincides with the density of the conditional r.v. $[W|Z=z]$, that is $p(*|z)$. Note that if $F(1) < \infty$, then $x \mapsto f(x)/F(1)$ is a pdf generating the same copula as f and $p(w|z)$ can be interpreted as a truncation model (Gagliardini and Gouriéroux, 2007). Actually, Gagliardini and Gouriéroux propose the f latent function in the context of non-parametric model estimation. Interestingly, it turns out that copula models are, in some sense, differentiable with respect to the functional parameter f , whereas they are not differentiable with respect to ϕ.

We shall now focus on the conditions for regular variation of f-generators. As opposed to F-generators, regular variation indices are only restricted at 0^+.

Proposition 4.1.8. *Let f be as in Proposition* 4.1.6 *and let $\alpha \in \mathbb{R}$. If $f \in \mathcal{R}_{0^+,\alpha}$, then $\alpha \geq -1$.*

Proof. Suppose $\alpha < -1$. Using Karamata's theorem (Corollary B.0.2), for some $x_0 > 0$ we have

$$F(x_0) - F(x) = \int_x^{x_0} t^\alpha \ell(t)\, dt \sim -\frac{x^{\alpha+1}}{\alpha+1}\ell(x) \ ,$$

as $x \to 0^+$. Since $\lim_{x\to 0^+} x^{\alpha+1}\ell(x) = \infty$, for Proposition B.0.6, f is not integrable, which leads to a contradiction.

Finally, let us remark that instances of regularly varying f's with index $\alpha \geq -1$ are possible. □

Regular variation indices of f and F are related through the following result.

Proposition 4.1.9. *Let f and F be as in Proposition* 4.1.6. *We have:*

1. *If $f \in \mathcal{R}_{0^+,\alpha}$, then $F \in \mathcal{R}_{0^+,\beta}$, where $\beta = \alpha + 1$.*
2. *If $f \in \mathcal{R}_{1^-,\alpha}$, then $F \in \mathcal{R}_{1^-,\beta}$, where $\beta = \min\{\alpha+1, 0\}$.*

Proof.

1. $F(0^+) = 0$, by construction. Hence, using L'Hôpital's rule,

$$\lim_{x\to 0^+} \frac{F(\lambda x)}{F(x)} = \lambda \lim_{x\to 0^+} \frac{f(\lambda x)}{f(x)} = \lambda^{\alpha+1} \ .$$

2. If $f \in \mathcal{R}_{1^-,\alpha}$, using Proposition B.0.5, there exists $x_0 \in (0,1)$ such that $f(x) = (1-x)^\alpha \ell(x)$ for all $x \in (x_0, 1)$.

   Suppose $\alpha < -1$. We have, for $x \in (x_0, 1)$,

$$F(x) = F(x_0) + \int_{x_0}^{x} (1-t)^\alpha \ell(t)\, dt$$

   and, by Karamata's theorem (Corollary B.0.2),

$$\int_{x_0}^{x} (1-t)^\alpha \ell(t)\, dt \sim -\frac{(1-x)^{\alpha+1}\ell(x)}{\alpha+1} \ ,$$

   as $x \to 1^-$.

Using the above results and the characterization of regular variation for continuously differentiable functions with normalized slowly varying part (Proposition $B.0.8$), which holds for an F-generator (it is monotonically increasing and we have Proposition $B.0.6$), it is possible to derive

$$\lim_{x\to 1^-} \frac{(1-x)f(x)}{F(x)} = \lim_{x\to 1^-} \frac{(1-x)^{\alpha+1}\ell(x)}{F(x_0) + \int_{x_0}^{x}(1-t)^{\alpha}\ell(t)\ dt} = -(\alpha+1)\ ,$$

where we have used that $\lim_{x\to 1^-}(1-x)^{\alpha+1}\ell(x) = \infty$, according to Remark $B.0.1$. Therefore, $\phi \in \mathcal{R}_{1^-,\alpha+1}$.

Now, suppose $\alpha > -1$. Using Karamata's theorem once more,

$$F(1) - F(x) = \int_x^1 (1-t)^{\alpha}\ell(t)\ dt \sim \frac{(1-x)^{\alpha+1}\ell(x)}{\alpha+1}\ ,$$

as $x \to 1^-$. Then, by Remark $B.0.1$, $\lim_{x\to 1^-}(1-x)^{\alpha+1}\ell(x) = 0$, which means $F(1) < \infty$. Therefore, $F \in \mathcal{R}_{1^-,0}$.

Finally, if $\alpha = -1$, we have either $F(1) < \infty$ or $F(1) = \infty$. In the former case, $F \in \mathcal{R}_{1^-,0}$, as in the $\alpha > -1$ case. The same holds for the latter. This can be shown using L'Hôpital's rule:

$$\lim_{x\to 0^+} \frac{F(1-\lambda x)}{F(1-x)} = \lambda \lim_{x\to 0^+} \frac{f(1-\lambda x)}{f(1-x)} = \lambda \cdot \lambda^{-1} = 1\ .$$

To sum up, if $\alpha < -1$, then $F \in \mathcal{R}_{1^-,\alpha+1}$, whereas if $\alpha \geq -1$, then $F \in \mathcal{R}_{1^-,0}$. This can be compactly expressed as the statement we wanted to prove.

□

Proposition 4.1.10 is useful to prove an intermediate tail dependence result (Charpentier and Segers, 2009) for Archimedean copulas, since $F(1)$ is in one-to-one correspondence with $\phi'(1)$.

Proposition 4.1.10. *Let $f \in \mathcal{R}_{1^-,\alpha}$ be an f-generator as in Proposition* 4.1.6 *and let F be its corresponding F-generator. We have:*

1. *If $\alpha < -1$, then $F(1) = \infty$.*
2. *If $\alpha > -1$, then $F(1) < \infty$.*
3. *If $\alpha = -1$, then $F(1)$ could be finite or infinite depending on f.*

Proof.

1. If $\alpha < -1$, then $F \in \mathcal{R}_{1^-,\beta}$, where $\beta = \alpha + 1 < 0$, for Proposition 4.1.9. Then, the result follows from Remark $B.0.1$.
2. This has been shown in the proof of Proposition 4.1.9.

3. Taking $f(x) = 1/(1-x)$ we get $F(1) = \infty$. To see an instance of the opposite case, consider $f(x) = (1-x)(\log(1-x))^2$ as $x \to 1^-$, i.e., there exists $x_0 > 0$ such that the equality holds if $x > x_0$. Then,

$$\begin{aligned} F(1) - F(x_0) &= \int_{x_0}^{1} \frac{1}{(1-t)[\log(1-t)]^2}\, dt \\ &= \left[\frac{-1}{\log(1-t)}\right]_{x_0}^{1} \\ &= \frac{-1}{\log(1-x_0)} < \infty\ . \end{aligned}$$

□

The following result summarizes how f and ϕ are related in terms of their regular variation behaviour.

Proposition 4.1.11. *Let ϕ and f be as in (4.1) and (4.6), respectively, and let $\alpha \in \mathbb{R}$.*

1. *If $f \in \mathcal{R}_{0^+,\alpha}$, then $\phi \in \mathcal{R}_{0^+,\beta}$, where $\beta = \min\{-\alpha, 0\}$.*
2. *If $f \in \mathcal{R}_{1^-,\alpha}$, then $\phi \in \mathcal{R}_{1^-,\beta}$, where $\beta = \max\{-\alpha, 1\}$.*

Proof. It is a direct consequence of Proposition 4.1.4 and Proposition 4.1.9, bearing in mind that $1 - \min\{\alpha + 1, 0\} = \max\{-\alpha, 1\}$. □

By applying Proposition 4.1.9 and Proposition 4.1.5, we can study the strictness of an Archimedean generator without referring to the intermediate construct F.

Proposition 4.1.12. *Let $f \in \mathcal{R}_{0^+,\alpha}$ be an f-generator as in Proposition* 4.1.6 *and let ϕ be its corresponding Archimedean generator. We have:*

1. *If $\alpha > 0$, then $\phi(0) = \infty$.*
2. *If $\alpha < 0$, then $\phi(0) < \infty$.*
3. *If $\alpha = 0$, then $\phi(0)$ could be finite or infinite depending on f.*

Proof. Just to clarify the case $\alpha = 0$, both $f(x) = 1$ and $f(x) = (1+\log x)\log x$, as $x \to 0^+$, are slowly varying at 0^+, but only the latter produces a non-strict ϕ, i.e., $\phi(0) < \infty$. □

Positivity and tail dependence

Even though the f-generator described in the previous section involves fairly weak and easy to implement requirements, there remain some details to work out. The foremost of them is, no doubt, the fact that the f in (4.6) may diverge both at 0 and 1, as Example 4.1.3 and Proposition 4.1.10 show. A piecewise definition of f that addresses regular variation at 0 and 1 would be feasible. The following alternative latent function, closely related to f, offers a more convenient solution, as we shall see.

Proposition 4.1.13 (*g*-**generator**). *Let $g : \mathbb{R} \to \mathbb{R}$ be continuous and define $f : (0,1) \to (0,\infty)$ as*

$$f(x) = \exp g\left[\sigma^{-1}(x)\right] \ , \tag{4.9}$$

where the logistic function $\sigma : \mathbb{R} \to (0,1)$ and its inverse σ^{-1} (known as the 'logit' function) are given by

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad \text{and} \quad \sigma^{-1}(x) = \log\left(\frac{x}{1-x}\right) \ ,$$

respectively. The function f is an f-generator provided that it is integrable on $[0,x]$, for all $x \in (0,1)$, in which case g is a g-generator.

Proof. The continuity of f is a consequence of that of g. The composition with the exponential function ensures the range of f is contained in $(0,\infty)$. □

Because two f's differing by a constant positive factor produce the same Archimedean copula, given a g-generator g and $K \in \mathbb{R}$, the map $x \mapsto g(x) + K$ is also a g-generator producing the same Archimedean copula as g.

The transformation (4.9) also works the other way round: an f-generator produces the g-generator

$$g(x) = \log f\left[\sigma(x)\right] \ . \tag{4.10}$$

In other words, g and f are in one to one correspondence. Hence, Proposition 4.1.7 applies in the case of g, too. Whenever f is the density function of some r.v. X, with corresponding CDF F, g can alternatively be expressed as the logarithm of the ratio of two densities:

$$g(x) = \log\left(\frac{\varsigma'(x)}{\sigma'(x)}\right) \ , \tag{4.11}$$

where $\varsigma(x) = F(\sigma(x)) = \mathbb{P}(X \leq \sigma(x)) = \mathbb{P}(\sigma^{-1}(X) \leq x)$ is the CDF of the r.v. $\sigma^{-1}(X)$ and σ is, of course, the CDF of the logistic distribution[10]. If $F(1) = \infty$, then f cannot represent a density. Nonetheless, equation (4.11) still holds, but ς would be a Lebesgue-Stieltjes function, rather than a CDF.

Figure 4.1 shows the plot of the logistic function and its derivative. As we can see, the shape of σ is *symmetric* in the sense that $\sigma(-x) = 1 - \sigma(x)$ and $\sigma'(x) = \sigma(x)(1-\sigma(x))$. Therefore, this map transforms in the same manner the x values to the left and to the right of zero. This feature is not exclusive of the logistic function; other non-skewed distribution functions, like the normal distribution Φ, could also be used. The advantage of using σ over, say, Φ is its simplicity and the fact that $\sigma'(x)$ can be directly computed in terms of $\sigma(x)$.

Figure 4.2 shows the extreme value distribution $S^{-1}(x) = \exp(-\exp(-x))$, which plays the role of σ in (Lambert, 2014), as discussed in Chapter 3, Section 3.1.2, and its density (derivative) $(S^{-1})'(x) = \exp[-(x + \exp(-x))]$. In this case, the transformation is skewed to the right of zero.

In what follows we will explore some regular variation properties of functions g and f that relate to tail dependence indices. The following result will also illustrate how the *symmetry* of σ allows to model the regular variation of f both at 0^+ and 1^- in a similar way, with a single function definition.

[10] For Proposition *C*.0.15, σ is actually the CDF of the r.v. $\sigma^{-1}(U)$, where U is a r.v. uniformly distributed in $[0,1]$. Therefore, (4.11) can be interpreted as the logarithm of the ratio of two logit-transformed r.v., namely $\sigma^{-1}(X)$ and $\sigma^{-1}(U)$.

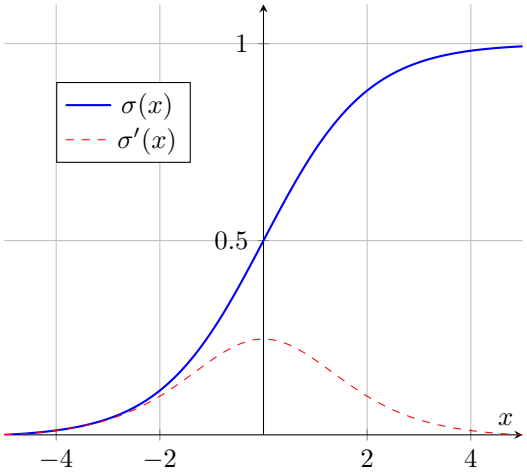


Figure 4.1: The logistic function σ and its derivative σ'

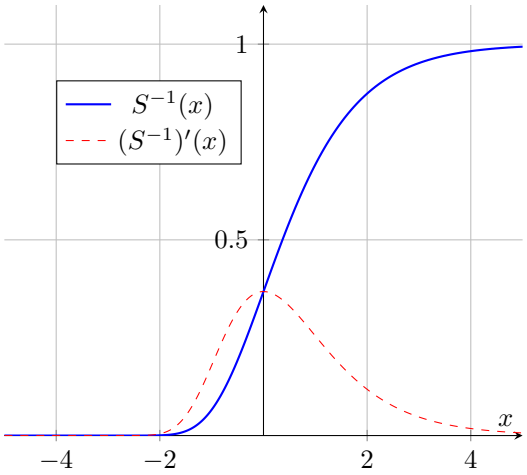


Figure 4.2: The extreme value distribution S^{-1} and its density $(S^{-1})'$

Proposition 4.1.14. *Let g and f be as in (4.9).*

1. *If $g \in \mathcal{R}^{+}_{-\infty,\alpha}$, then $f \in \mathcal{R}_{0^+,\alpha}$.*
2. *If $g \in \mathcal{R}^{+}_{\infty,\alpha}$, then $f \in \mathcal{R}_{1^-,-\alpha}$.*

Proof. Let $g \in \mathcal{R}^{+}_{-\infty,\alpha}$. From Proposition $B.0.10$, we have

$$g(x) = \alpha x + \ell(x) \ ,$$

as $x \to -\infty$, where $\ell \in S^{+}_{-\infty}$. This means that

$$f(x) = x^{\alpha}\tilde{\ell}(x) \ , \tag{4.12}$$

as $x \to 0^+$, where $\tilde{\ell}(x) = (1-x)^{-\alpha} \exp \ell\left(\sigma^{-1}(x)\right)$. Now, according to Proposition B.0.5, we just have to check that $\tilde{\ell} \in \mathcal{S}_{0^+}$. Given $\lambda > 0$,

$$\lim_{x\to 0^+} \frac{\tilde{\ell}(\lambda x)}{\tilde{\ell}(x)} = \exp\left[\lim_{x\to 0^+} \ell\left(\sigma^{-1}(\lambda x)\right) - \ell\left(\sigma^{-1}(x)\right)\right] \ . \tag{4.13}$$

An easy calculation shows that

$$\sigma^{-1}(\lambda x) = \sigma^{-1}(x) + \mu\left(\sigma^{-1}(x)\right) \ , \tag{4.14}$$

with

$$\mu(x) = \log\lambda - \log\left(1 + (1-\lambda)e^{x}\right) \ , \tag{4.15}$$

so

$$\lim_{x\to 0^+} \ell\left(\sigma^{-1}(\lambda x)\right) - \ell\left(\sigma^{-1}(x)\right) = \lim_{x\to -\infty} \ell\left(x + \mu(x)\right) - \ell(x) \ . \tag{4.16}$$

The right-hand side of last expression resembles that of Definition $B.0.8$, except for the fact that μ is not constant. To account for this fact, we note that slow variation implies uniform convergence (Proposition $B.0.11$). Indeed, given $\delta > 0$, since $\mu(x) \to \log\lambda$ as $x \to -\infty$, there exists x_1 such that $|\mu(x) - \log\lambda| < \delta$ whenever $x < x_1$. Furthermore, fixing $I = [\log\lambda - \delta, \log\lambda + \delta]$, we know that for all $\epsilon > 0$ there exists x_2 such that $|\ell(x+\mu_0) - \ell(x)| < \epsilon$ if $x < x_2$ and $\mu_0 \in I$. Hence, taking $x_3 < \min\{x_1, x_2\}$, we get $\mu(x) \in I$ and, consequently, $|\ell\left(x + \mu(x)\right) - \ell(x)| < \epsilon$, if $x < x_3$.

The steps to demonstrate the case $g \in \mathcal{R}^{+}_{\infty,\alpha}$ are the same, with some minor and obvious changes:

1. Instead of (4.12), we have

   $$f(x) = (1-x)^{-\alpha}\tilde{\ell}(x) \ ,$$

   as $x \to 1^-$, where $\tilde{\ell}(x) = x^{\alpha} \exp \ell\left(\sigma^{-1}(x)\right)$ and $\ell \in S^{+}_{\infty}$.

2. Instead of (4.13),

   $$\lim_{x\to 0^+} \frac{\tilde{\ell}(1-\lambda x)}{\tilde{\ell}(1-x)} = \exp\left[\lim_{x\to 0^+} \ell\left(\sigma^{-1}(1-\lambda x)\right) - \ell\left(\sigma^{-1}(1-x)\right)\right] \ .$$

3. Instead of (4.14) and (4.15), using $\sigma^{-1}(*) = -\sigma^{-1}(1-*)$ as a shortcut, we arrive at $\sigma^{-1}(1-\lambda x) = \sigma^{-1}(1-x) + \mu\left(\sigma^{-1}(1-x)\right)$, for $\mu(x) = \log\left(1+(1-\lambda)e^{-x}\right) - \log\lambda$, with $\mu(x) \to -\log\lambda$ as $x \to \infty$.

4. Instead of (4.16),

$$\lim_{x\to 0^+} \ell\left(\sigma^{-1}(1-\lambda x)\right) - \ell\left(\sigma^{-1}(1-x)\right) = \lim_{x\to\infty} \ell\left(x+\mu(x)\right) - \ell(x)\ . \quad (4.17)$$

5. Finally, the uniform convergence theorem, along with some derivations that involve ϵ and δ, similar to the ones made earlier, allows one to conclude that the limit (4.17) is zero.

□

The previous result holds even if g does not produce an integrable f. However, in order to be a legitimate g-generator, g must satisfy the following constraint.

Proposition 4.1.15. *Let g be g-generator, as described in Proposition* 4.1.13*, and let $\alpha \in \mathbb{R}$. If $g \in \mathcal{R}^+_{-\infty,\alpha}$, necessarily $\alpha \geq -1$.*

Proof. It is a direct consequence of Proposition 4.1.8 and Proposition 4.1.14. □

The following result relates the additive regular variation of g and the regular variation of ϕ.

Proposition 4.1.16. *Let g be a g-generator and let ϕ be its corresponding Archimedean generator. Letting $\alpha \in \mathbb{R}$, we have:*

1. *If $g \in \mathcal{R}^+_{-\infty,\alpha}$, then $\phi \in \mathcal{R}_{0^+,\beta}$, where $\beta = \min\{-\alpha, 0\}$.*
2. *If $g \in \mathcal{R}^+_{\infty,\alpha}$, then $\phi \in \mathcal{R}_{1^-,\beta}$, where $\beta = \max\{\alpha, 1\}$.*

Proof. It directly follows from Proposition 4.1.11 and Proposition 4.1.14. □

The strictness property of an Archimedean generator is related to the additive regular variation of g.

Proposition 4.1.17. *Let $g \in \mathcal{R}^+_{-\infty,\alpha}$ be a g-generator as in Proposition* 4.1.13 *and let ϕ be its corresponding Archimedean generator. We have:*

1. *If $\alpha > 0$, then $\phi(0) = \infty$.*
2. *If $\alpha < 0$, then $\phi(0) < \infty$.*
3. *If $\alpha = 0$, then $\phi(0)$ is finite or infinite depending on g.*

Proof. This is a consequence of Proposition 4.1.12, thanks to Proposition 4.1.14. For examples that illustrate opposite instances of the limiting case $\alpha = 0$, look at those of Proposition 4.1.12 and apply (4.10). □

We also have some control over f producing a finite measure or not by adjusting the additive regular variation index of g.

Proposition 4.1.18. *Let $g \in \mathcal{R}^+_{\infty,\alpha}$ be a g-generator as in Proposition* 4.1.13 *and let F be its corresponding F-generator, via (4.9) and (4.6). We have:*

1. *If* $\alpha > 1$, *then* $F(1) = \infty$.
2. *If* $\alpha < 1$, *then* $F(1) < \infty$.
3. *If* $\alpha = 1$, *then* $F(1)$ *is finite or infinite depending on* g.

Proof. It follows from Proposition 4.1.10, via Proposition 4.1.14. □

Approximations of the generator function in terms of splines

Cubic splines and B-splines have been used to approximate a sufficiently simple latent function like the g we have just introduced. In (Lambert, 2014), reviewed in Chapter 3, Section 3.1.2, it was the derivative of a certain generator which was expressed as a linear combination of B-splines over a bounded domain. In that work, little or no attention was paid to the asymptotic properties of the approximation as $|x| \to \infty$. Two major issues are worth noting at this respect:

- In (Lambert, 2014), not every polynomial continuation provides a valid construction.
- In (Lambert, 2014), even if a polynomial continuation applies as $x \to -\infty$, the function to be approximated could grow so fast that it would not be possible to represent it with a polynomial, always resulting in lower tail independence, as Proposition 3.1.5 and Figure 3.4 show.

The work by Hernández-Lobato and Suárez relies on the approximation of g with natural cubic splines. The parametrization used in our work is not the original proposal in (Hernández-Lobato and Suárez, 2011), but actually a simpler and more convenient one in which the slopes of the spline as $|x| \to \infty$ become explicit parameters. For a complete explanation on this matter, we refer the reader to Appendix A.

Proposition 4.1.19 (**Natural spline** g**-generator).** *Let* $\Delta = \{a = \xi_1 < \xi_2 < \ldots < \xi_n = b\}$ *be an* n*-point partition of the interval* $[a, b] \subset \mathbb{R}$. *Let* $N_1, N_2, \ldots, N_n$ *be the natural splines basis associated with* Δ, *as described in Appendix A, Section A.2.1. Let* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^T \in \mathbb{R}^n$ *and define* $g : \mathbb{R} \to \mathbb{R}$ *as*

$$g(x) = \sum_{i=1}^{n} \theta_i N_i(x) \ , \tag{4.18}$$

where we assume the convention that the basis functions N_i*'s extend outside* $[a, b]$ *by means of a linear extrapolation*[11]. *If* $\theta_1 > -1$, *then (4.18) is a* g*-generator.*

Proof. Remember that $g'(a) = \theta_1$ in our parametrization. Since g is a natural spline, with a linear continuation for $x < a$, we may assume that, for all $x < a$, $g(x) = \theta_1 x + K$ for some $K \in \mathbb{R}$, thus $g \in \mathcal{R}^+_{-\infty,\theta_1}$. Then, Proposition 4.1.15 only allows $\theta_1 \geq -1$. Finally, we simply need to check the limiting case $\theta_1 = -1$ is not allowed either. Using (4.9),

$$f(x) = \exp\left(-\log\left(\frac{x}{1-x}\right) + K\right) = e^K \left(\frac{1-x}{x}\right) \ ,$$

which is not integrable near 0. □

[11] See (A.14) in the appendix.

Proposition 4.1.19 explicitly introduces, for the first time in this chapter, the underlying parameter vector $\boldsymbol{\theta}$ of Hernández-Lobato and Suárez's model. If we were to optimize those parameters, the copula model would not be *identifiable*, which means that two distinct parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can give rise to the same copula. This poses some theoretical objections to the use of maximum likelihood estimation, for example. The following result explains why (4.18) suffers from this identifiability issue and, at the same time, gives us the key to solve it.

Proposition 4.1.20. *Let g be a natural spline g-generator as in (4.18). For every $k = 2, 3, \ldots, n-1$, there exists a unique natural spline g-generator $g_{-k} : \mathbb{R} \to \mathbb{R}$ having a zero k-th coordinate with respect to the basis in (4.18) and producing the same Archimedean copula as g.*

Proof. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^T$ be the coordinates of the natural spline g in (4.18) with respect to the natural spline basis $N_1, N_2, \ldots, N_n$. Similarly, let $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \ldots, \vartheta_{n+2})^T$ be the coordinates of g with respect to the cubic B-splines $B_1, B_2, \ldots, B_{n+2}$. For Proposition A.2.1, building a spline by adding a constant $K \in \mathbb{R}$ to all the coordinates ϑ_i's is equivalent to adding the same constant K to the original spline. More precisely, letting g_K have coordinates $\vartheta_i + K$ with respect to the cubic B-splines, we have

$$g_K(x) = \sum_{i=1}^{n+2} (\vartheta_i + K) B_i(x) = K + \sum_{i=1}^{n+2} \vartheta_i B_i(x) = g(x) + K \ , \tag{4.19}$$

for all $x \in [a, b]$.

Now, looking at Proposition A.2.7, with the caveat that the notation employed to refer to $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ is different, we can easily see that adding a constant K to each component of $\boldsymbol{\vartheta}$ is equivalent to adding the same constant K to each component of $(\theta_2, \theta_3, \ldots, \theta_{n-1})^T$, i.e., all the coordinates of $\boldsymbol{\theta}$ except for the first and last ones. Therefore, (4.19) can equivalently be stated as

$$\begin{aligned} g_K(x) &= \theta_1 N_1(x) + \theta_n N_n(x) + \sum_{i=2}^{n-1} (\theta_i + K) N_i(x) \\ &= K + \sum_{i=1}^{n} \theta_i N_i(x) \\ &= g(x) + K \ , \end{aligned} \tag{4.20}$$

for all $x \in [a, b]$. Note that, since g and g_K are natural splines, equation (4.20) holds for $x \in \mathbb{R}$.

Finally, let k be any index in $\{2, 3, \ldots, n-1\}$ and consider $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_n)^T$, where $\tilde{\theta}_1 = \theta_1$, $\tilde{\theta}_n = \theta_n$, $\tilde{\theta}_k = 0$ and $\tilde{\theta}_i = \theta_i - \theta_k$ if $i \neq k$. Defining $g_{-\theta_k} : \mathbb{R} \to \mathbb{R}$ as

$$g_{-\theta_k}(x) = \sum_{i=1}^{n} \tilde{\theta}_i N_i(x) = \sum_{\substack{i=1 \\ i \neq k}}^{n} \tilde{\theta}_i N_i(x) \tag{4.21}$$

and taking $K = -\theta_k$ in (4.20), we have that

$$g_{-\theta_k}(x) = g(x) - \theta_k \ , \tag{4.22}$$

for all $x \in \mathbb{R}$. All in all, $g_{-\theta_k}$ in (4.21) produces the same Archimedean copula as g, because of (4.22), while having a zero k-th coordinate. □

According to Proposition 4.1.20, to make (4.18) identifiable it suffices to remove the k-th basis function, N_k, for some $k \in \{2, 3, \ldots, n-1\}$, and find our candidate natural spline

$$g(x) = \sum_{\substack{i=1 \\ i \neq k}}^{n} \theta_i N_i(x) \ , \tag{4.23}$$

by optimizing the reduced parameter vector $(\theta_1, \theta_2, \ldots, \theta_{k-1}, \theta_{k+1}, \theta_{k+2}, \ldots, \theta_n)^T \in \mathbb{R}^{n-1}$. A secondary but convenient consequence of modelling (4.23) instead of (4.18) is that we need to optimize one less parameter, saving us some computations.

The natural spline specification of g is additively regularly varying at $\pm\infty$ by construction.

Remark 4.1.2. Let g be a natural spline g-generator as described in Proposition 4.1.19.[12] Remember that $g'(a) = \theta_1$ and $g'(b) = \theta_n$. On the other hand, because of the natural spline constraints, $g''(a) = g''(b) = 0$, the linear extrapolation

$$g(x) = \begin{cases} g(a) + \theta_1(x-a), & \text{if } x < a \\ g(b) + \theta_n(x-b), & \text{if } x > b \end{cases} \ , \tag{4.24}$$

implies $g \in \mathcal{C}^2(\mathbb{R}) \cap \mathcal{R}^+_{-\infty,\theta_1} \cap \mathcal{R}^+_{\infty,\theta_n}$.

As a result, the parameters θ_1 and θ_n determine the regularly varying behaviour of ϕ.

Proposition 4.1.21. *Let g be a natural spline g-generator as described in Proposition* 4.1.19*, with θ_1 and θ_n as its first and last coordinates, respectively, and let ϕ be its corresponding Archimedean generator. We have:*

1. *$\phi \in \mathcal{R}_{0^+,\beta}$, where $\beta = \min\{-\theta_1, 0\}$.*
2. *$\phi \in \mathcal{R}_{1^-,\beta}$, where $\beta = \max\{\theta_n, 1\}$.*

Moreover, the θ_1 parameter is related to the strictness of the resulting Archimedean generator ϕ, while θ_n determines the finiteness of $F(1)$.

Proposition 4.1.22. *Let g be a natural spline g-generator as described in Proposition* 4.1.19*, with θ_1 as its first coordinate, and let ϕ be its corresponding Archimedean generator. We have:*

1. *If $\theta_1 \geq 0$, then $\phi(0) = \infty$.*
2. *If $\theta_1 < 0$, then $\phi(0) < \infty$.*

Proof. The case $\theta_1 \neq 0$ easily follows from Proposition 4.1.17, via Remark 4.1.2, so we just have to check that $\theta_1 = 0$ produces a strict Archimedean generator. In this case, $g(x) = K \in \mathbb{R}$ for all $x < a$. Then, $f(x) = e^K$ and $F(x) \propto x$ if $x < \sigma(a)$. Finally, $\phi(x) \propto -\log x$, thus $\phi(0) = \infty$. □

[12] The general construction subsumes the identifiable version (4.23).

Proposition 4.1.23. *Let g be a natural spline g-generator as described in Proposition* 4.1.19*, with θ_n as its last coordinate, and let F be its corresponding F-generator, via (4.9) and (4.6). We have:*

1. *If $\theta_n \geq 1$, then $F(1) = \infty$.*
2. *If $\theta_n < 1$, then $F(1) < \infty$.*

Proof. Proposition 4.1.18 is sufficient to determine $F(1)$ from θ_n whenever $\theta_n \neq 1$. If $\theta_n = 1$, we have $g(x) = x + K$, for all $x > b$ and some $K \in \mathbb{R}$. Hence, using (4.10), $f(x) \propto x/(1-x)$ for x near 1 and thus $F(1) = \infty$. □

The left and right linear regions, $(-\infty, a)$ and (b, ∞), respectively, are related to tail events, whereas the interval $[a, b]$, which represents the non-linear region of the spline, is related to non-tail events, as the reader might have guessed by now from Remark 4.1.2 and Proposition 2.4.6. A major contribution of (Hernández-Lobato and Suárez, 2011) is taking advantage of the linear regions (4.24) to model tail dependencies. Actually, nothing prevents us from using higher order polynomial approximations, but the scarcity of data to model tail dependencies discourages from employing complex models outside $[a, b]$. Remembering Domingos's words:

> "... strong false assumptions can be better than weak true ones, because a learner with the latter needs more data to avoid overfitting."
>
> (Domingos, 2012)

In this sense, the linear extrapolation (Remark 4.1.2) is the simplest (and so the strongest assumption) among all possible additively regularly varying continuations, with a constant slowly varying part.

A key aspect in the design of the latent function g is the location of the spline knots ($\{\xi_i\}_{i=1}^n$). There are a number of reasonable ways to address this task, but none of them seems clearly superior to the others. In contexts where the function to be approximated is graphically linked to data, like in a regression setting, reasonable guesses can be made; it is even possible to optimize the location of the B-spline knots (Mao and Zhao, 2003). In contexts like ours, where the latent function g is related to data only in an indirect manner that involves several (integral) transformations, the problem is more difficult to solve. (Vandenhende and Lambert, 2005, Lambert, 2007, 2014) recommend employing a large number of equidistant knots.

(Hernández-Lobato and Suárez, 2011) propose a different approach based on Theorem 4.1.1 that suggests knots can be placed on empirical grounds. In any case, Hernández-Lobato and Suárez propose to employ also a large number of knots ($n = 20$, for example).

Definition 4.1.1 (Kendall knots). Let $\mathcal{D} = \{(U_i, V_i)\}_{i=1}^N$ be a sample contained in $[0,1]^2$. From $\mathcal{D}$, we can obtain an empirical estimate $C_{\mathcal{D}}$ of the underlying copula and, then, an empirical estimate of the corresponding Kendall distribution. Let $\hat{W}$ be the r.v. corresponding to the empirical Kendall distribution arising from the pseudo-sample $\{W_i\}_{i=1}^N$, where $W_i = C_{\mathcal{D}}(U_i, V_i)$. We define the Kendall knots $\{q_i\}_{i=1}^n$ as the quantiles

$$q_i = \mathcal{Q}_{\sigma^{-1}(\hat{W})}(p_i) \ ,$$

where $p_1 = \epsilon$ and $p_n = 1 - \epsilon$, for some $\epsilon \gtrsim 0$ (typically $\epsilon = 0.01$), and

$$p_i = p_1 + \frac{p_n - p_1}{n-1}(i-1) \ ,$$

for $i = 2, 3, \ldots, n-1$.

The motivation for Definition 4.1.1 can be better understood by looking at equation (4.8) in Theorem 4.1.1, which can be rewritten in terms of g as

$$p(w,z) = \frac{\exp g\left[\sigma^{-1}(w)\right]}{F(z)} \mathbb{1}_{[0,z]}(w) = \frac{\exp g\left[\sigma^{-1}(w)\right]}{\int_{-\infty}^{\sigma^{-1}(z)} \exp g(t)\ d\sigma(t)} \mathbb{1}_{[0,z]}(w) \ , \quad (4.25)$$

where $d\sigma(t) = \sigma'(t)\ dt$. If we want to estimate the joint density of (W, Z), where W is the Kendall r.v. and Z is one of the uniform marginal r.v., from a sample $\{W_i, Z_i\}_{i=1}^N$ using model (4.25), placing the spline knots of g according to Definition 4.1.1 can be a reasonable choice. By definition, Kendall knots accumulate the same proportion of points $\{\sigma^{-1}(W_i)\}_{i=1}^N$ between any two consecutive knots q_k and q_{k+1}, leaving proportions ϵ and $1-\epsilon$ to the left of q_1 and to the right of q_n, respectively. Since the argument of g in the numerator in (4.25) is $\sigma^{-1}(w)$, Kendall knots evenly distribute sample instances $\sigma^{-1}(W_i)$'s among spline pieces. Hence, all parameters θ_i's in (4.23) should be evenly validated. Note, however, that this interpretation does not take into account the integral in the denominator of (4.25).

The knot placement proposed in Hernández-Lobato and Suárez perform reasonably well in most situations. However, they tend to misplace the rightmost knot before the linear regime in the upper tail is reached, which does not allow one to provide good estimates for $g'(b) = \theta_n$. The reason for this *bias* lies in the empirical distribution of $\{\sigma^{-1}(W_i)\}_{i=1}^N$, on which Kendall knots depend. Since each W_i is bounded from above by either U_i or V_i, which are uniformly distributed, the distribution of $\sigma^{-1}(\hat{W})$ is asymmetric with respect to $\sigma^{-1}(1/2) = 0$; roughly speaking, there are usually (many) more negative values than positives ones, typically resulting in $0 < q_n < -q_1$. This is even more marked for small values of Kendall's tau, since, according to Proposition 2.3.3,

$$\mathbb{E}(W) = \frac{1 + \tau(C)}{4} \ ,$$

where W represents the true Kendall r.v. associated with the copula estimate C. We shall present some examples of samples $\{\sigma^{-1}(W_i)\}_{i=1}^N$ that illustrate this phenomenon in Chapter 5.

If we do not feel drawn to the empirical Kendall knots, but rather prefer to *stick* to the non-informative equidistant knots, as many others, there are basically two ways to do that. Hernández-Lobato and Suárez's analogue approach to Lambert's equidistant knots in $[S(\epsilon), S(1-\epsilon)]$ can be obtained by replacing S with σ^{-1}. This means that the knots are equidistant in the domain of g (of g', in (Lambert, 2014)).

Definition 4.1.2 (Uniform knots). We define the uniform knots $\{q_i\}_{i=1}^n$ as the quantiles

$$q_i = \mathcal{Q}_{\sigma^{-1}(U)}(p_i) = \sigma^{-1}(p_i) \ ,$$

where U is a uniformly distributed r.v. in $[0, 1]$ and the values of the p_i's are defined as in Definition 4.1.1.

The linearity assumptions entailed by the use of natural splines attempt to reduce the complexity of the model where data instances are scarce. If linear approximations are valid when $|x| \gg 0$ and if g is sufficiently well-behaved, we can argue that the curvature of g, which represents the complexity of the model, is not *uniform* over the non-linear region, but rather increases as $|x|$ approaches 0. On the other hand, one could argue that knots should be placed where the curvature of g is bigger. Combining this two ideas, we may conclude that knots should accumulate near 0 and that is what Definition 4.1.2 suggests. Observe that uniform knots are not equidistant in the domain of g (the whole $\mathbb{R}$), but their images through σ, the p_i's, are uniform in the domain of f: the interval $(0, 1)$.

4.1.2 Relation to dependence measures

In this section we will analyse the capabilities of Hernández-Lobato and Suárez's technique to model certain dependence measures, some of which have already been introduced in Chapter 2.

Concordance

There are several ways to express the Kendall's tau of Hernández-Lobato and Suárez's estimate. Unfortunately all of them involve two different generators. Most notably, we have

$$\begin{aligned} \tau(C_\phi) &= 1 - 4\int_0^1 F(x)\phi(x)\ dx \\ &= 4\int_0^1 xf(x)\phi(x)\ dx - 1 \end{aligned}, \tag{4.26}$$

where f, F and ϕ are the generators introduced in Section 4.1.1. The latter expression is computationally more accurate than the former, because evaluating F instead of f requires an additional approximation step (see Appendix D). In either case, there does not seem to be any strong limitation to the representation of Kendall's tau in this model, provided that we use enough knots.

Tail dependence

The following proposition is useful to understand why the modelling of the tail dependence is one of the strongest points of (Hernández-Lobato and Suárez, 2011).

Proposition 4.1.24. *Let g be a natural spline g-generator as described in Proposition* 4.1.19*, with θ_1 and θ_n as its first and last coordinates, respectively, and let ϕ be its corresponding Archimedean generator. The Archimedean copula generated by ϕ, C_ϕ, has lower and upper tail indices*

$$\lambda_L(C_\phi) = \begin{cases} 2^{-1/\theta_1}, & \text{if } \theta_1 > 0 \\ 0, & \text{if } \theta_1 \leq 0 \end{cases}$$

and

$$\lambda_U(C_\phi) = \begin{cases} 2 - 2^{1/\theta_n}, & \text{if } \theta_n > 1 \\ 0, & \text{if } \theta_n \leq 1 \end{cases},$$

respectively.

Proof. It easily follows from Proposition 4.1.21 and Proposition 2.4.6. □

As we can see, no upper or lower tail index is out of reach. Interestingly, neither the upper nor the lower tail index are in one-to-one correspondence with their associated slopes. Namely, $\lambda_L(C_\phi) = 0$ and $\lambda_U(C_\phi) = 0$ for infinite values of θ_1 and θ_n, respectively. Therefore, it is natural to wonder what differences, if any, there are among the different tail dependencies. A partial answer to this question is given by the following tail dependence indices, introduced by Ledford and Tawn according to (Hua and Joe, 2013).

Definition 4.1.3 (**Ledford & Tawn's indices).** Let C be a bivariate copula. The Ledford & Tawn's lower and upper tail indices are defined as

$$\begin{aligned} \eta_L(C) &= \lim_{u\to 0^+} \frac{\log u}{\log C(u,u)} \\ \eta_U(C) &= \lim_{u\to 0^+} \frac{\log u}{\log \hat{C}(u,u)} \end{aligned},$$

respectively, provided the limits exist, where $\hat{C}$ represents the survival copula of C. Both indices are related through the identity $\eta_U(C) = \eta_L(\hat{C})$.

Ledford & Tawn's indices are in the interval $[0, 1]$. A pivotal value is $1/2$, which corresponds to the independence copula. Values larger than $1/2$ correspond thus to some positive dependence; smaller values, to negative dependence. Ledford & Tawn's indices are a measure of *intermediate* tail dependence, in the sense that they are not sensitive to different values of the standard tail indices, but they might distinguish between different cases of standard tail independence.

Remark 4.1.3. If C is a bivariate copula such that $\lambda_L(C) > 0$, then

$$\eta_L(C) = \lim_{u\to 0^+} \frac{\log u}{\log C(u,u)} = \left[1 + \lim_{u\to 0^+} \frac{\log\left(\frac{C(u,u)}{u}\right)}{\log u}\right]^{-1} = 1 \ .$$

Similarly, if $\lambda_U(C) > 0$, then $\eta_U(C) = 1$.

The following result taken from (Charpentier and Segers, 2009) summarizes all possibles values of Ledford & Tawn's upper tail index for Archimedean copulas, under the standard regular variation hypothesis.

Theorem 4.1.2. *Let $\phi \in \mathbf{\Phi}$ and let C_ϕ be the Archimedean copula generated by ϕ. If $\phi \in \mathcal{R}_{1^-,1}$, then*

$$\eta_U(C_\phi) = \begin{cases} 1/2, & \text{if } \phi'(1) < 0 \\ 1, & \text{if } \phi'(1) = 0 \end{cases} .$$

The previous result shows that regular variation limits to a great extent the application of Ledford & Tawn's upper tail index, with only two possible values. Hua and Joe present an example of Archimedean copula family $\{C_\alpha\}_{\alpha\in(0,1)}$, expressed in terms of Laplace transforms, such that $\eta_U(C_\alpha) = 1/(1+\alpha)$ and, hence, $\eta_U(C_\alpha)$ takes all values in $(1/2, 1)$.

The Archimedean copulas obtained using Hernández-Lobato and Suárez's technique are regularly varying. Therefore, we can apply Theorem 4.1.2 to obtain the following corollary.

Corollary 4.1.2. *Let g be a natural spline g-generator as described in Proposition* 4.1.19*, with θ_n as its last coordinate, and let ϕ be its corresponding Archimedean generator. The Archimedean copula generated by ϕ, C_ϕ, has Ledford & Tawn's upper tail index*

$$\eta_U(C_\phi) = \begin{cases} 1/2, & \text{if } \theta_n < 1 \\ 1, & \text{if } \theta_n \geq 1 \end{cases} .$$

Proof. It is a consequence of Theorem 4.1.2 and Proposition 4.1.23, noting $\phi'(1) = -1/F(1)$. □

Corollary 4.1.2 shows that Hernández-Lobato and Suárez's technique is as good as they come as refers to upper tail dependence modelling. Even more interesting is the combination $\lambda_U(C_\phi) = 0$ and $\eta_U(C_\phi) = 1$, corresponding to $\theta_n = 1$. According to (Charpentier and Segers, 2009), this is a *rare* case, known as near asymptotic dependence, which does not usually arise in customary parametric models.

The next result, actually a particular case of a more general one in (Charpentier and Segers, 2009), contrasts with the limitations imposed by Theorem 4.1.2.

Theorem 4.1.3. *Let $\phi \in \mathbf{\Phi}_\infty$ and let C_ϕ be the Archimedean copula generated by ϕ. If $\phi \in \mathcal{S}_{0^+}$ and $\psi \in \mathcal{R}_{\infty,\alpha}$, for some $\alpha \in \mathbb{R}$, where*

$$\psi(x) = \left[\frac{d}{dx} \left\{ \log \phi^{-1} \right\} (x) \right]^{-1} ,$$

then $\alpha \leq 1$ and $\eta_L(C_\phi) = 2^{\alpha-1}$.

For Archimedean copulas, hence, Ledford & Tawn's lower tail index can take any value in $(0, 1]$, even under the regular variation assumption. However, in the representation given by Hernández-Lobato and Suárez, it is not possible to span the whole range of values.

Proposition 4.1.25. *Let g be a natural spline g-generator as described in Proposition* 4.1.19*, with θ_1 as its first coordinate, and let ϕ be its corresponding Archimedean generator. The Archimedean copula generated by ϕ, C_ϕ, has Ledford & Tawn's lower tail index*

$$\eta_L(C_\phi) = \begin{cases} 0, & \text{if } \theta_1 < 0 \\ 1/2, & \text{if } \theta_1 = 0 \\ 1, & \text{if } \theta_1 > 0 \end{cases} .$$

Proof. The case $\theta_1 > 0$ follows from Proposition 4.1.24 and Remark 4.1.3. The case $\theta_1 < 0$ implies $\phi(0) < \infty$, for Proposition 4.1.22. Therefore, $C_\phi(u, u) = 0$ for a sufficiently small $u > 0$ and the result follows by Definition 4.1.3. Finally, the case $\theta_1 = 0$ can be checked either by a simple calculation or realizing that a constant g produces the independence copula ($\theta_1 = 0$ means g is constant as $x \to -\infty$). □

4.1.3 Estimation and smoothing

In this section we address the problem of estimating the parameters in (4.23), avoiding overfitting. We will also provide useful implementation details, introduced by Hernández-Lobato and Suárez. Even though we basically follow their recommendations, we introduce some modifications that improve the effectiveness of the method. The reader is referred to Section D.1.4 for a detailed description of these improvements and the insights that motivated them.

Maximum likelihood estimation

Maximum likelihood is the estimation method proposed in (Hernández-Lobato and Suárez, 2011). Rearranging factors in the Archimedean copula density (4.4), using (4.5), we get

$$c_\phi(u,v) = f(C_\phi(u,v))\frac{F(C_\phi(u,v))}{F(u)F(v)} \ , \tag{4.27}$$

where f and F are the corresponding generators. Now, including explicitly the dependence on $\boldsymbol{\theta}$, the vector of parameters in terms of which g is expressed, we get

$$\log c_{\boldsymbol{\theta}}(u,v) = g_{\boldsymbol{\theta}}\left\{\sigma^{-1}[C_{\boldsymbol{\theta}}(u,v)]\right\} + \log\frac{F_{\boldsymbol{\theta}}[C_{\boldsymbol{\theta}}(u,v)]}{F_{\boldsymbol{\theta}}(u)F_{\boldsymbol{\theta}}(v)} \ , \tag{4.28}$$

where we have used (4.9). The log-likelihood of $\boldsymbol{\theta}$ given the data set $\mathcal{D} = \{(U_i,V_i)\}_{i=1}^N$ is

$$\log\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) = \sum_{i=1}^N \log c_{\boldsymbol{\theta}}(U_i,V_i) \ . \tag{4.29}$$

It is important to remark that $\boldsymbol{\theta}$ actually refers to the reduced parameter vector in (4.23), which was introduced to make the model identifiable. We shall write $\boldsymbol{\theta} \in \mathbb{R}^{n-1}$ to emphasize the missing k-th natural spline basis function in the expansion of g.

The original proposal (4.29) in (Hernández-Lobato and Suárez, 2011) is not the only loss function one can optimize to estimate $\boldsymbol{\theta}$. We can also consider the density (4.25) corresponding to the r.v. (W,Z) described in Theorem 4.1.1, where W follows the Kendall distribution of the Archimedean copula and Z is any of the uniform margins. The log-density of the model is

$$\log p_{\boldsymbol{\theta}}(w,z) = g_{\boldsymbol{\theta}}\left(\sigma^{-1}(w)\right) - \log F_{\boldsymbol{\theta}}(z) \ , \tag{4.30}$$

where the indicator function $\mathbb{1}_{[0,z]}$ has been omitted on the grounds that, in practice, we shall always have $0 \le w \le z$ for empirical realizations of (W,Z). Observe that, because the Kendall distribution produces, under certain hypotheses, an Archimedean copula, the loss function

$$\log\mathcal{L}(\boldsymbol{\theta}|\tilde{\mathcal{D}}) = \sum_{i=1}^N \log p_{\boldsymbol{\theta}}(W_i,Z_i) \ , \tag{4.31}$$

where $\tilde{\mathcal{D}} = \{(W_i,Z_i)\}_{i=1}^N$, for $Z_i = U_i$ (or equivalently, $Z_i = V_i$) and W_i as defined in Definition 4.1.1, could also be used. As a matter of fact, we have seen a similar approach, based on the Kendall distribution, in (Vandenhende

and Lambert, 2005). Fairly speaking, Vandenhende and Lambert had arguably no choice, since they could not rely on a copula density function.

One of the weak points of using the loss function (4.31) is the estimation of the pseudo sample $\{W_i\}_{i=1}^N$. Replacing $C_{\boldsymbol{\theta}}(U_i, V_i)$ with W_i entails an information loss, since we need to use an empirical estimate of $C_{\boldsymbol{\theta}}$ subject to sample errors. However, the accuracy of empirical estimates grows for large sample sizes and, moreover, there are a number of sophisticated non-parametric techniques that can be explored, too.

The strength of (4.31) over (4.29) is the simplicity of (4.30), as compared with (4.28). The former involves just two latent functions ($g_{\boldsymbol{\theta}}$ and $F_{\boldsymbol{\theta}}$) and, besides, is devoid of complex compositions involving functions that depend on $\boldsymbol{\theta}$. We can take advantage of these features during optimization, as we shall see in Appendix D.

Smoothing formulae

The actual optimization problem in (Hernández-Lobato and Suárez, 2011) incorporates a term $E(\boldsymbol{\theta})$ that penalizes the curvature of the $g_{\boldsymbol{\theta}}$ function, which, as discussed above, represents the complexity of the model:

$$E(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left[g''_{\boldsymbol{\theta}}(x)\right]^2 \, dx = \int_a^b \left[g''_{\boldsymbol{\theta}}(x)\right]^2 \, dx \ , \tag{4.32}$$

since, of course, $g''_{\boldsymbol{\theta}}(x) = 0$ for all $x \notin (a, b)$.

In Appendix A, the curvature (4.32) of a spline was expressed in terms of the coordinates with respect to the B-splines basis, rather than the natural basis, but a similar expression can be found. If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{k-1}, \theta_{k+1}, \theta_{k+2}, \ldots, \theta_n) \in \mathbb{R}^{n-1}$, consider the augmented coordinates vector $\bar{\boldsymbol{\theta}} = (\theta_1, \theta_2, \ldots, \theta_{k-1}, 0, \theta_{k+1}, \theta_{k+2}, \ldots, \theta_n) \in \mathbb{R}^n$ corresponding to the full expansion of $g_{\boldsymbol{\theta}}$. On the other hand, consider $\Omega \in \mathcal{M}_{n\times n}(\mathbb{R})$ with elements

$$\Omega_{ij} = \int_a^b N_i''(x) N_j''(x) \, dx \tag{4.33}$$

and the reduced matrix $\Omega^* \in \mathcal{M}_{(n-1)\times(n-1)}(\mathbb{R})$ formed by removing the k-th row and column in Ω.[13] Then, the curvature becomes as simple as

$$E(\boldsymbol{\theta}) = \bar{\boldsymbol{\theta}}^T \Omega \bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^T \Omega^* \boldsymbol{\theta} \ . \tag{4.34}$$

The curvature (4.32) plays a similar role to (3.26), the norm of the r-th order differences of $\boldsymbol{\theta}$, in (Lambert, 2014). In our quest for a genuine semiparametric method, we reckon that the curvature (4.32) is a more meaningful measure than (3.26), because it is rooted in a physical concept that involves the functional parameter g. For instance, the norm of the differences vector is not sensitive to knot locations, which are essential in the description of splines, whereas the curvature definitely takes them into account.

Optimization procedure

The optimization problem to solve in (Hernández-Lobato and Suárez, 2011) is

$$\arg\max_{\boldsymbol{\theta}\in\Theta} \left\{\log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) - \lambda E(\boldsymbol{\theta})\right\} \ , \tag{4.35}$$

[13] Without any row or column permutation.

with parameter space

$$\Theta = \{(\theta_1, \theta_2, \ldots, \theta_{k-1}, \theta_{k+1}, \theta_{k+2}, \ldots, \theta_n) \in \mathbb{R}^{n-1} \ : \ \theta_1 \geq 0\} \tag{4.36}$$

and penalty factor $\lambda \geq 0$. The original loss function in (Hernández-Lobato and Suárez, 2011) is (4.29), but we can use (4.31) instead.

The constraint $\theta_1 \geq 0$ in (4.36), Hernández-Lobato and Suárez's original proposal, ensures $\phi_{\boldsymbol{\theta}}(0) = \infty$, according to Proposition 4.1.22. It can be relaxed so as to allow for non-strict Archimedean generators, requiring $\theta_1 > -1$ instead, as stated in Proposition 4.1.19. Strictly speaking, the original natural spline parametrization in (Hernández-Lobato and Suárez, 2011) expresses Θ in a slightly more complicated manner, since $g'_{\boldsymbol{\theta}}(a)$ does not correspond to θ_1, but to a linear combination of the full set of parameters.

Penalty λ is determined via cross-validation. Namely, Hernández-Lobato and Suárez recommend using a 10-fold cross-validation grid search. This means that the log-likelihood in (4.35) uses only 90% of the available data for training, while the remaining 10% is devoted to validation. Once obtained a cross-validated penalty λ^*, the final parameter vector $\boldsymbol{\theta}$ is estimated by plugging $\lambda = \lambda^*$ into (4.35) and making use of the whole dataset $\mathcal{D}$. Since the magnitude of $\log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$ grows as the sample size N does, the penalized log-likelihood used as loss function after cross-validation is not the same as the one used during cross-validation to obtain λ^*. This might reduce the effectiveness of the K-fold cross-validation, specially if we employ fewer folds, like $K = 4$ or $K = 3$, in which case the size of datasets during validation account for just 75% or 66% of all available data. Therefore, we propose to adjust (4.36) to *normalize* the log-likelihood according to the size of the dataset employed at each step and use instead

$$\arg\max_{\boldsymbol{\theta} \in \Theta} \left\{ \frac{1}{N} \log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) - \lambda E(\boldsymbol{\theta}) \right\} . \tag{4.37}$$

The cross-validation proposed in (Hernández-Lobato and Suárez, 2011) may be very time-consuming, specially if we want to test a large number of candidate λ's. Hernández-Lobato and Suárez simplify the grid search by considering log-spaced candidate λ's; namely, the (natural) logarithms of the candidates are the integers $1, 2, 3, 4$. Note that, in the N-normalized version (4.37), these particular integer values are not supposed to work; instead we propose to try $\lambda = 10^k$ for several $k \in \mathbb{Z}$.

If we work under tight time constraints and cannot afford to perform a cross-validation step, we may resort to a Bayesian penalty approach, even if our main interest is not Bayesian statistics. We can easily adapt the proposal in (Lambert, 2014) to our problem by replacing the r-th order difference Gramian matrix P in (3.34) with Ω^*. Taking hyper-parameters $a = b = 1$, as recommended in (Lambert, 2014), we can alternatively solve the problem

$$\arg\max_{\boldsymbol{\theta} \in \Theta} \left\{ \log \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) - \left(1 + \frac{\operatorname{rank}(\Omega^*)}{2}\right) \log\left(1 + \frac{1}{2}\boldsymbol{\theta}^T \Omega^* \boldsymbol{\theta}\right) \right\} . \tag{4.38}$$

The rationale behind the prior distribution proposed in (Lambert, 2007, 2014) also applies when penalizing the curvature. Since Ω^* is positive semi-definite[14], it is the Gramian matrix of a set of vectors, that is, $\Omega^* = A^T A$, for

[14] By construction, $\boldsymbol{\theta}^T \Omega^* \boldsymbol{\theta} \geq 0$ for all $\boldsymbol{\theta} \in \mathbb{R}^{n-1}$.

some real matrix A such that $\text{rank}(\Omega^*) = \text{rank}(A)$. Therefore, $E(\boldsymbol{\theta})$ can be interpreted as $E(\boldsymbol{\theta}) = \|A\boldsymbol{\theta}\|^2$ and $A\boldsymbol{\theta}$ is the curvature counterpart of the r-th order differences vector $\Delta^r_{n-1}\boldsymbol{\theta}$ in (Lambert, 2014).

Problems (4.37) and (4.38) cannot be analytically solved, so we must resort to numerical methods. We refer the reader to Appendix D for further details on implementation.

4.2 Semiparametric Archimedean copula family

In this section we will present a new approach to conditional copula modelling based on the bivariate copula modelling technique in (Hernández-Lobato and Suárez, 2011). We must praise Hernández-Lobato and Suárez for the introduction of such a convenient method, whose light constraints make the conditional extension a feasible goal. On the other hand, the nice approximating features semiparametric Archimedean copulas (SPAC) have are also present in the newer SPAC family members.

4.2.1 The construction of a bivariate latent function

Much like in (Lambert, 2014), our goal is to find a suitable family of SPAC generators $\{\phi_w\}_{w\in[0,1]}$ and then consider the conditional copula family $\{C(*,*|w) = C_{\phi_w}\}_{w\in[0,1]}$. We build our new proposal both upon the concept of g-generator and the spline approximation techniques. That being said, we will skip the motivation steps we took to arrive to the final operating version of the g-generator, based on natural cubic splines.

We are looking for a bivariate function $g : \mathbb{R} \times [0,1] \to \mathbb{R}$ such that, fixed a covariate value $w \in [0,1]$, $x \mapsto g(x,w)$ is a natural spline g-generator as defined in Proposition 4.1.19. Let $N_1, N_2, \ldots, N_n$ be the usual natural spline basis in Proposition 4.1.19, associated with a partition $\Delta_x = \{\xi_i^x\}_{i=1}^n$, where $a = \xi_1^x < \xi_2^x < \ldots < \xi_n^x = b$. For now, consider

$$g(x,w) = \sum_{i=1}^{n} \theta_i(w) N_i(x) \ , \tag{4.39}$$

where $\theta_i : [0,1] \to \mathbb{R}$, for $i = 1, 2, \ldots, n$, and $\theta_1(w) > -1$, for all $w \in [0,1]$.[15] From (4.39), we can obtain ϕ_w in three steps. Firstly, construct the f-generator

$$f_w(x) = f(x,w) = \exp g(\sigma^{-1}(x), w) \ .$$

Then, integrate to obtain the F-generator

$$F_w(x) = F(x,w) = \int_0^x f_w(t)\ dt \tag{4.40}$$

and finally the Archimedean generator

$$\phi_w(x) = \phi(x,w) = \int_x^1 \frac{1}{F_w(t)}\ dt \ .$$

Since we are proficient at approximating with splines, it comes as no surprise the following definition (Lyche and Schumaker, 1975, Lyche and Morken, 2002).

[15] Remember from Proposition 4.1.19 that the first coordinate θ_1 of a natural spline, with respect to the basis $N_1, N_2, \ldots, N_n$, must satisfy $\theta_1 > -1$.

Definition 4.2.1 (**Tensor product spline g-generator**). Let $\Delta_w = \{\xi_i^w\}_{i=1}^M$ be a partition of the covariate space $[0,1]$, i.e., $0 = \xi_1^w < \xi_2^w < \ldots < \xi_M^w = 1$. Let $B_1, B_2, \ldots, B_m$, with $m = M + 2$, be the cubic B-splines associated with the partition Δ_w. We can take each coordinate function θ_i in (4.39) as

$$\theta_i(w) = \sum_{j=1}^{m} \theta_{ij} B_j(w) \ ,$$

for a parameter matrix $\Theta = (\theta_{ij}) \in \mathcal{M}_{n\times m}(\mathbb{R})$ such that $\theta_{1j} > -1$, for $j = 1, 2, \ldots, m$. Then, we can define the tensor product spline g-generator

$$g_\Theta(x, w) = \sum_{i=1}^{n} \sum_{j=1}^{m} \theta_{ij} N_i(x) B_j(w) \ . \tag{4.41}$$

Just like in Proposition 4.1.19, we need $\theta_1(w) > -1$ for all $w \in [0,1]$. For Proposition A.2.1, it suffices to take $\theta_{1j} > -1$ for $j = 1, 2, \ldots, m$. Similarly, if we want to ensure each ϕ_w is strict, i.e., $\phi_w(0) = \infty$, we may consider $\theta_{1j} \geq 0$ for $j = 1, 2, \ldots, m$.

We have no reliable heuristic to place the ξ_i^w knots over the covariate space. We recommend employing equidistant knots in $[0,1]$, i.e., $\xi_i^w = (i-1)/(M-1)$, for $i = 1, 2, \ldots, M$. As for the knots in the x variable, ξ_i^x, note that Kendall knots (Definition 4.1.1) do not make sense in the conditional case, where we typically have a single data instance (U_i, V_i, W_i) for each conditional copula: we have no repeated values in the covariates data set $\mathcal{D}_W = \{W_i\}_{i=1}^N$. We propose to use uniform knots (Definition 4.1.2), which, in fact, were our first option for the non-conditional case (not Hernández-Lobato and Suárez's).

Fixed $x \in \mathbb{R}$, we have that $g_w = w \mapsto g_\Theta(x, w)$ is a cubic spline. This alone represents an improvement with respect to (Lambert, 2014), where the counterpart of g_w is not exactly a spline, but a squared spline. On the other hand, just like in (Lambert, 2014), the construction (4.41) can be easily extended to include more covariates, just adding more B-spline dimensions and indices.[16] Note also that we could have used a cubic natural spline basis for the w-dimension, too, but the natural spline constraints do not make so much sense in the covariate space.

There is an identifiability issue in Definition 4.2.1, just like in the original definition of natural spline g-generator (Proposition 4.1.19). Given $\Theta = (\theta_{ij}) \in \mathcal{M}_{n\times m}(\mathbb{R})$, consider $\tilde{\Theta} = (\tilde{\theta}_{ij}) \in \mathcal{M}_{n\times m}(\mathbb{R})$ with

$$\tilde{\theta}_{ij} = \begin{cases} \theta_{ij} + \alpha_j, & \text{if } i \neq 1 \text{ and } i \neq n \\ \theta_{ij}, & \text{if } i = 1 \text{ or } i = n \end{cases} \ ,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$. For Proposition 4.1.20, we have

$$g_{\tilde{\Theta}}(x, w) = g_\Theta(x, w) + \sum_{j=1}^{m} \alpha_j B_j(w) \ ,$$

which means Θ and $\tilde{\Theta}$ produce the same Archimedean copula family. We must avoid such an identifiability issue both for theoretical and practical reasons.

[16] Obviously, at the expense of raising computational costs.

We can attain this by simply removing N_{i^*}, provided that $i^* \neq 1$ and $i^* \neq n$, which is equivalent indeed to setting $\alpha_j = -\theta_{i^*j}$, for all $j = 1, 2, \ldots, m$, so that $\tilde{\theta}_{i^*j} = 0$. Hence, once chosen i^*, we shall write

$$g_\Theta(x, w) = \sum_{\substack{i=1 \\ i \neq i^*}}^{n} \sum_{j=1}^{m} \theta_{ij} N_i(x) B_j(w) \ , \tag{4.42}$$

for a reduced parameter matrix

$$\Theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \ldots & \theta_{1m} \\ \theta_{21} & \theta_{22} & \ldots & \theta_{2m} \\ \vdots & \vdots & & \vdots \\ \theta_{(i^*-1)1} & \theta_{(i^*-1)2} & \ldots & \theta_{(i^*-1)m} \\ \theta_{(i^*+1)1} & \theta_{(i^*+1)2} & \ldots & \theta_{(i^*+1)m} \\ \vdots & \vdots & & \vdots \\ \theta_{n1} & \theta_{n2} & \ldots & \theta_{nm} \end{bmatrix} \ ,$$

where the i^*-th row from the original matrix specification is missing, for a total of $(n-1)m$ parameters. Nonetheless, we shall denote the original specification by $\bar{\Theta} = (\theta_{ij})$, assuming $\theta_{i^*j} = 0$ for all $j = 1, 2, \ldots, m$, that is

$$\bar{\Theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \ldots & \theta_{1m} \\ \theta_{21} & \theta_{22} & \ldots & \theta_{2m} \\ \vdots & \vdots & & \vdots \\ \theta_{(i^*-1)1} & \theta_{(i^*-1)2} & \ldots & \theta_{(i^*-1)m} \\ 0 & 0 & \ldots & 0 \\ \theta_{(i^*+1)1} & \theta_{(i^*+1)2} & \ldots & \theta_{(i^*+1)m} \\ \vdots & \vdots & & \vdots \\ \theta_{n1} & \theta_{n2} & \ldots & \theta_{nm} \end{bmatrix} \ . \tag{4.43}$$

4.2.2 Estimation and smoothing

As previously seen in this work, the estimation and smoothing of the θ_{ij}'s admit both a Bayesian and a non-Bayesian approach. The latter can be solely stated in terms of the likelihood function, to which a penalty term may optionally be added. In the former approach, though, estimation and smoothing are the two sides of the same coin and cannot be clearly separated: the smoothing penalty is incorporated into the model as prior knowledge, which is essential in the Bayesian approach to estimation. In this work, though, we shall content ourselves with a shallow application of Bayesian statistics, using the Bayesian prior in (Lambert, 2007, 2014) as a mere regularization term. We shall not explore posterior distribution sampling.

Let $c_\Theta(*, *|w)$ be the density of the Archimedean copula family generated by (4.42) and consider the dataset $\mathcal{D} = \{(U_i, V_i, W_i)\}_{i=1}^N$. As usual, the likelihood function of the copula family, for both the non-Bayesian and the Bayesian approach, is

$$\mathcal{L}(\Theta|\mathcal{D}) = \prod_{i=1}^{N} c_\Theta(U_i, V_i|W_i) \tag{4.44}$$

and, of course,

$$\log c_\Theta(u, v|w) = g_\Theta\left\{\sigma^{-1}[C_\Theta(u, v|w)], w\right\} + \log \frac{F_\Theta[C_\Theta(u, v|w), w]}{F_\Theta(u, w)F_\Theta(v, w)} , \quad (4.45)$$

where

$$C_\Theta(u, v|w) = \phi_\Theta^{-1}(\phi_\Theta(u, w) + \phi_\Theta(v, w), w) .$$

In either case, we need to devise a way to measure the overfitting of the model (4.42). Remember that Lambert proposed a measure based on the vector differences which we did not find full and satisfactory for a semiparametric model (for instance, vector differences do not take into account the location of the spline knots). Instead, we propose to use an energy measure that extends the notion of curvature employed in (Hernández-Lobato and Suárez, 2011) and reviewed in Section 4.1.3 and Appendix A: the *thin plate* energy measure (Wood and Augustin, 2002).

Definition 4.2.2 (Thin plate energy measure). Let g be as in (4.39). We define the thin plate energy of g as

$$E(g) = \int_a^b \int_0^1 \left[\left(\frac{\partial^2 g}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 g}{\partial x \partial w}\right)^2 + \left(\frac{\partial^2 g}{\partial w^2}\right)^2\right](x, w)\, dx\, dw . \quad (4.46)$$

Expression (4.46) becomes specially tractable when g is a tensor product spline as in Definition 4.2.1.

Proposition 4.2.1. *Let g_Θ as in (4.42). We have*

$$E(\Theta) = E(g_\Theta) = (\operatorname{vec}\bar{\Theta})^T \cdot \Omega \cdot \operatorname{vec}\bar{\Theta} ,$$

for some matrix $\Omega \in \mathcal{M}_{nm\times nm}(\mathbb{R})$, where $\bar{\Theta}$ is as described in (4.43).

Proof. Denote by $N_i^{(p)}$ and $B_j^{(q)}$ the p-th and q-th derivatives of N_i and B_j, respectively, assuming $N_i^{(0)} = N_i$ and $B_j^{(0)} = B_j$. Define matrices $N^p = (N_{ik}^p)_{i,k=1}^n$ and $B^q = (B_{jl}^q)_{j,l=1}^m$, where

$$\left.\begin{aligned} N_{ik}^p &= \int_a^b N_i^{(p)}(x) \cdot N_k^{(p)}(x)\, dx \\ B_{jl}^q &= \int_0^1 B_j^{(q)}(w) \cdot B_l^{(q)}(w)\, dw \end{aligned}\right. ,$$

for $p = 0, 1, 2$ and $q = 0, 1, 2$. An easy calculation shows

$$\int_a^b \int_0^1 \left[\frac{\partial^{p+q} g_\Theta}{\partial x^p \partial w^q}(x, w)\right]^2 dx\, dw = \sum_{\substack{i=1\\k=1}}^{n} \sum_{\substack{j=1\\l=1}}^{m} \theta_{ij}\theta_{kl} N_{ik}^p B_{jl}^q .$$

Now, since

$$(B^q \otimes N^p)_{n(j-1)+i, n(l-1)+k} = N_{ik}^p B_{jl}^q ,$$

and

$$\left.\begin{aligned} (\operatorname{vec}\bar{\Theta})_{n(j-1)+i} &= \theta_{ij} \\ (\operatorname{vec}\bar{\Theta})_{n(l-1)+k} &= \theta_{kl} \end{aligned}\right. ,$$

we have

$$\sum_{\substack{i=1\\k=1}}^{n}\sum_{\substack{j=1\\l=1}}^{m}\theta_{ij}\theta_{kl}N_{ik}^{p}B_{jl}^{q}=\sum_{\mu=1}^{nm}\sum_{\nu=1}^{nm}(\operatorname{vec}\bar{\Theta})_{\mu}(B^{q}\otimes N^{p})_{\mu\nu}(\operatorname{vec}\bar{\Theta})_{\nu}\ .$$

Therefore,

$$\int_{a}^{b}\int_{0}^{1}\left[\frac{\partial^{p+q}g_{\Theta}}{\partial x^{p}\partial w^{q}}(x,w)\right]^{2}\ dx\ dw=(\operatorname{vec}\bar{\Theta})^{T}\cdot(B^{q}\otimes N^{p})\cdot\operatorname{vec}\bar{\Theta}\ .$$

The result follows from last equation, taking

$$\Omega=\sum_{p+q=2}\frac{2}{p!\times q!}(B^{q}\otimes N^{p})\ .$$

□

Proposition 4.2.1 can be restated in terms of the reduced matrix Θ. Using some basic results in (Wood and Augustin, 2002), we have

$$E(\Theta)=\left(\operatorname{vec}\bar{\Theta}^{T}\right)^{T}\cdot\Omega\cdot\operatorname{vec}\bar{\Theta}^{T}\ , \tag{4.47}$$

for a new Ω matrix

$$\Omega=\sum_{p+q=2}\frac{2}{p!\times q!}(N^{p}\otimes B^{q})\ . \tag{4.48}$$

Finally, it easily follows from (4.47) that

$$E(\Theta)=\left(\operatorname{vec}\Theta^{T}\right)^{T}\cdot\Omega^{*}\cdot\operatorname{vec}\Theta^{T}\ ,$$

where Ω^{*} is a reduced version of the Ω in (4.48) without the rows and columns ranging from $(i^{*}-1)m-1$ to $i^{*}m$.

Our non-Bayesian proposal is analogous to the non-conditional case (4.37): for a cross-validated $\lambda\geq 0$, solve the average penalized log-likelihood optimization problem

$$\arg\max_{\Theta\in\mathbf{\Theta}}\left\{\frac{1}{N}\log\mathcal{L}(\Theta|\mathcal{D})-\lambda E(\Theta)\right\}\ , \tag{4.49}$$

with parameter space

$$\mathbf{\Theta}=\{\Theta\in\mathcal{M}_{(n-1)\times m}(\mathbb{R}):\theta_{1j}>-1\text{ for }j=1,2,\ldots,m\}\ .$$

Similarly, for the Bayesian approach, the optimization problem becomes

$$\arg\max_{\Theta\in\mathbf{\Theta}}\left\{\log\mathcal{L}(\Theta|\mathcal{D})-\left(1+\frac{\operatorname{rank}(\Omega^{*})}{2}\right)\log\left(1+\frac{E(\Theta)}{2}\right)\right\}\ . \tag{4.50}$$

Further details on how to implement the tensor product spline construction and the estimation and smoothing procedures are gathered in Appendix D.

Model tuning considerations

Unlike (Lambert, 2014), we believe that the construction (4.42) can be effective even for relatively small values of n and m. The current trend in statistical modelling favours employing a large number of parameters and then introducing some sort of penalty to reduce overfitting. We can feasibly apply this philosophy in the non-conditional version (Hernández-Lobato and Suárez, 2011), but not in the conditional one, unless we apply some sort of additive parameter configuration such as (Lambert, 2014).

Model tuning may become a daunting task if we have to test for several values of n and m (plus the penalty λ, if we opt for the non-Bayesian smoothing). We propose setting n equal to $M = m - 2$ and considering only small values for n, say, $3 \leq n \leq 7$, for a total of $(n-1)(n+2)$ parameters. Then, if we choose to apply any type of smoothing, we would recommend to take the largest admissible value for n, like $n = 7$; if we take $\lambda = 0$, then we might prefer to focus on selecting the optimum n.

Chapter 5

Simulation study

> – ¿Cómo puedes estar seguro?
> – Tenemos nuestros Cerebros electrónicos, Noys; calculadoras mucho más exactas que cualquier otra que se haya podido inventar en cualquier Realidad. Podemos analizar las posibles realidades y evaluar las ventajas entre miles y miles de variables.
>
> *El fin de la eternidad*
> Isaac Asimov

This chapter is devoted to presenting graphical evidence of the effectiveness of the Archimedean methods studied in Chapter 4, specially of our proposal (Section 4.2). We will exclusively focus on synthetic data obtained from some well-known copula families, both Archimedean and non-Archimedean, all of which were introduced in Section 2.4.

5.1 Bivariate copula simulations

In this section we will present further evidence of the capability of (Hernández-Lobato and Suárez, 2011), while testing our own implementation: an essential step in order to extend Hernández-Lobato and Suárez's technique to conditional copulas.

All the code used to perform the following experiments is written in Julia (Bezanson et al., 2012). Despite the shortage of external packages for Julia, which made us implement some basic but important parts, including a splines library, one feels really comfortable programming in Julia, thanks to its friendly syntax, interactive environment and powerful functional programming capabilities. Julia's built-in parallelism was also useful to speed up computations.

5.1.1 Simulation settings and procedure

We will perform simulations for a number of copula families: Clayton, Frank and Gumbel (Archimedean families in Table 2.1) and the Gaussian family (an example of elliptical family). We will also check different values of Kendall's tau for each copula family, namely $\tau = 0.2$, $\tau = 0.5$ and $\tau = 0.8$. Given a value

τ_0 for Kendall's tau, for the families we are interested in, we can always find a copula C such that $\tau(C) = \tau_0$. Namely, for the Clayton, Gumbel and Gaussian families, the parameters we are looking for are

$$\theta_{\mathrm{Cl}}(\tau) = \frac{2\tau}{1-\tau} \ , \ \theta_{\mathrm{Gu}}(\tau) = \frac{1}{1-\tau} \ , \ \theta_{\mathrm{Ga}}(\tau) = \sin\left(\frac{\pi\tau}{2}\right) \ ,$$

respectively. Remember that the parameter of the bivariate Gaussian copula is a linear correlation coefficient. For the Frank family there is no closed-form expression for θ as a function of τ, but we can resort to a root-finding algorithm to get an approximation.

Given a copula family and a value for τ, we shall simulate $S = 20$ datasets $\mathcal{D}_i = \left\{\left(U_k^{(i)}, V_k^{(i)}\right)\right\}_{k=1}^{N}$ of $N = 1000$ bivariate points from that copula. Then, from each $\mathcal{D}_i$ we estimate a natural spline g-generator $\tilde{g}_i$, for $i = 1, 2, \ldots, S$, from which the rest of relevant functions, namely $\tilde{f}_i$, $\tilde{F}_i$, $\tilde{\phi}_i$ and $\tilde{\phi}_i^{-1}$, can be obtained. Only five knots were used to model the g-generator, with no smoothing penalty. Strictness was enforced in all $\tilde{\phi}_i$'s.

Bearing in mind, from (4.26), that the density of the Kendall's distribution of an Archimedean copula C_ϕ is

$$K'_{C_\phi}(x) = f(x)\phi(x) \ , \tag{5.1}$$

we also define, for $i = 1, 2, \ldots, S$, the estimate of (5.1) from $\mathcal{D}_i$:

$$k_i(x) = f_i(x)\phi_i(x) \ .$$

Finally, to make results clearer, we can apply the logistic function σ so as to evaluate (5.1) (and its estimates) over the whole $\mathbb{R}$. Let W be a unidimensional r.v. distributed according to K_{C_ϕ}. We have that $\sigma^{-1}(W)$ is distributed according to

$$H(x) = \mathbb{P}(\sigma^{-1}(W) \leq x) = \mathbb{P}(W \leq \sigma(x)) = K_{C_\phi}(\sigma(x)) \ ,$$

supported over $\mathbb{R}$, with density

$$h(x) = H'(x) = K'_{C_\phi}(\sigma(x)) \cdot \sigma'(x) \ . \tag{5.2}$$

Then, estimates of (5.2) can be obtained for each dataset $\mathcal{D}_i$:

$$h_i(x) = k_i(\sigma(x)) \cdot \sigma'(x) \ .$$

With all the above information, two different kinds of visual proofs are provided. One type aims at evaluating how successfully the estimation process fitted the original Archimedean g-generators g and Kendall densities K'_{C_ϕ}.[1] Look, for instance, at Figure 5.1. Each subplot comprises two halves that share the same x-axis. The upper half shows:

- The average $\bar{g}$ of the g-generator estimates $\tilde{g}_i$. If the original copula was Archimedean with g-generator g, in order to make a visually fair comparison between $\bar{g}$ and g, we make sure $\bar{g}(0) = g(0)$. To do this, we may just redefine $\bar{g}$ as

$$\bar{g}(x) = \frac{1}{S}\sum_{i=1}^{S}[\tilde{g}_i(x) - \tilde{g}_i(0) + g(0)] \ .$$

[1] Remember that the Kendall distribution uniquely defines an Archimedean copula (Theorem 2.4.2).

Note that this redefinition only adds a constant to the previous average function and, hence, produces the same Archimedean copula.

- The maximum and minimum of $\tilde{g}_i(x)$, for x taking values on a fine grid, that is,
$$\min_{1\leq i\leq S}\{\tilde{g}_i(x)\}\ ,\ \max_{1\leq i\leq S}\{\tilde{g}_i(x)\}\ .$$

- The 0.05 and 0.95 sample quantiles of $\tilde{g}_i(x)$, that is, the values $q_{0.05}(x)$ and $q_{0.95}(x)$ that are greater or equal than $\tilde{g}_i(x)$ for 5% and 95%, respectively, of the simulations $i = 1, 2, \ldots, S$.

On the other hand, the lower half shows:

- The average $\bar{h}$ of the Kendall density estimates $\tilde{h}_i$:
$$\bar{h}(x) = \frac{1}{S}\sum_{i=1}^{S}\tilde{h}_i(x)\ .$$

- The maximum and minimum of $\tilde{h}_i(x)$, for x taking values on a fine grid, that is,
$$\min_{1\leq i\leq S}\{\tilde{h}_i(x)\}\ ,\ \max_{1\leq i\leq S}\{\tilde{h}_i(x)\}\ .$$

- The 0.05 and 0.95 sample quantiles of $\tilde{h}_i(x)$.

- An histogram of the empirical estimate of $\sigma^{-1}(W)$ for the $\mathcal{D}_1$ dataset.

Both the upper and lower halves are crossed by five vertical dashed lines that mark the location of the natural spline knots. Uniform knots (Definition 4.1.2) were used.

For each family and Kendall's tau combination, the second type of plot, like Figure 5.5, compares the sample $\mathcal{D}_1$ and the level curves of the original copula density[2] with a new sample $\tilde{\mathcal{D}}_1$ of size $N = 1000$ drawn from the estimate $\tilde{g}_1$ and the level curves of the copula density estimate. The Archimedean copula estimates were sampled using a generic multivariate quantile-based algorithm (Embrechts et al., 2003); algorithms for LT-Achimedean copulas were also used (Hofert, 2008) to sample the original copulas.

5.1.2 Comments

The results obtained in Figures 5.1, 5.2, 5.3, 5.4 demonstrate a high average accuracy when estimating the g-generators and Kendall densities. Generator averages almost coincide with the original ones, specially in the $[-5, 5]$ range. In general, fluctuations in the estimation of the g-generator seem to be *bigger* than Kendall densities show, but both errors are actually equivalent.

Generator slopes when $|x| \to \infty$ (and thus tail dependence indices) are reasonably well captured in all the Archimedean examples. Estimation variances grow in both tails, but this comes as no surprise, given the shortage of eligible data instances to model tail dependence. On the other hand, the Gaussian

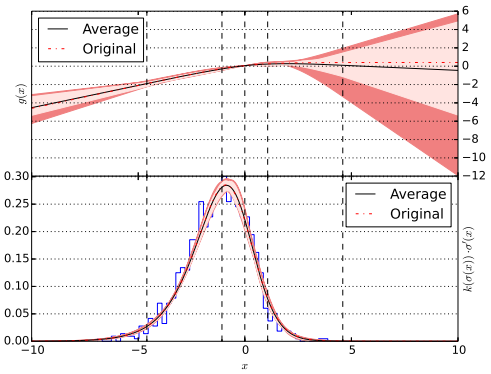
[2] Remember that the density of an Archimedean copula is given by (4.27).

copula has no tail dependence at all.[3] By looking at the generator slope as $x \to -\infty$, we can see that lower tail dependence is erroneously inferred in all cases for the Gaussian copula. Notwithstanding, upper tail dependence remains quite small, specially in the $\tau = 0.2$ and $\tau = 0.5$ cases, taking into account that slopes not exceeding 1 imply upper tail independence (Proposition 4.1.24).

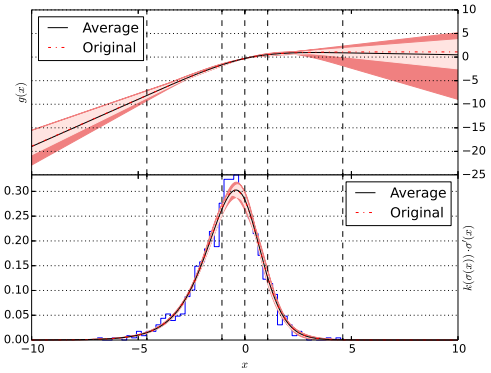
The simulations in Figures 5.5, 5.6, 5.7, 5.8 confirm the effectiveness of Hernández-Lobato and Suárez's technique. The density level curves are remarkably well captured, even in the Gaussian case, considering the size of $\mathcal{D}_1$: $N = 1000$. The sample $\tilde{\mathcal{D}}_1$ drawn from the learned copula and the training sample $\mathcal{D}_1$ are alike; Kendall's tau is reasonably well captured.

It is worth mentioning that, according to Figure 5.4, even if the Gaussian copula is not Archimedean, the estimation process seems to have successfully found some *sort* of projection onto the Archimedean copula space on the grounds of the Kendall distribution, since the estimated Kendall density fits the empirical sample fairly well.

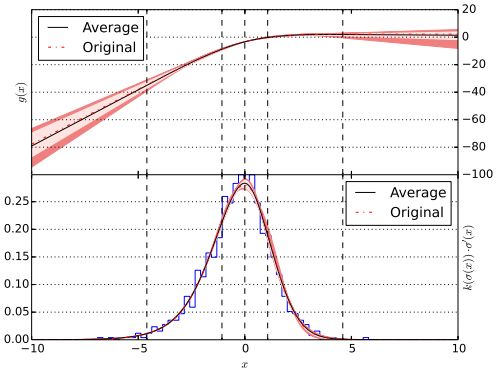
[3] Remember that the Gaussian copula is radially symmetric, so the upper and lower tails coincide.



(a) $\tau = 0.2$

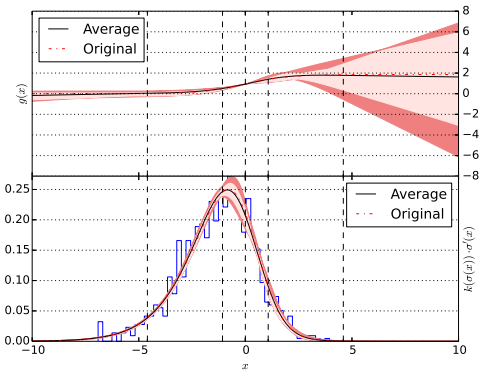


(b) $\tau = 0.5$

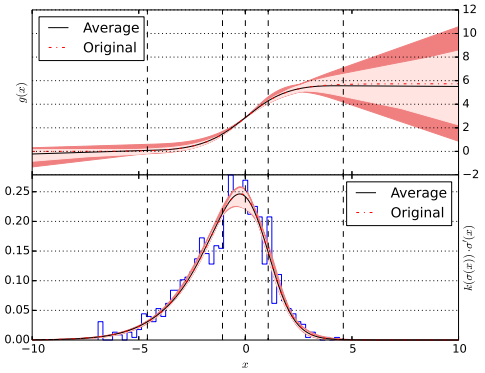


(c) $\tau = 0.8$

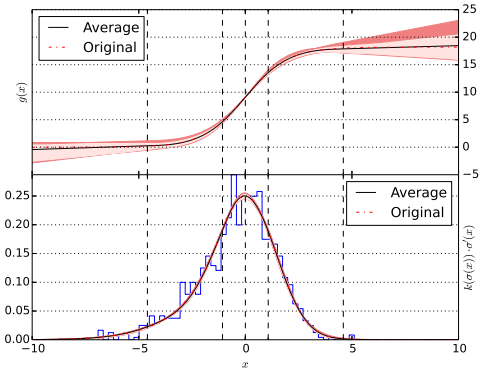
Figure 5.1: Clayton copula g-generator and Kendall density



(a) $\tau = 0.2$

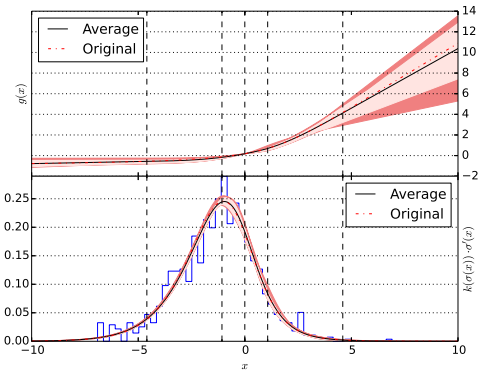


(b) $\tau = 0.5$

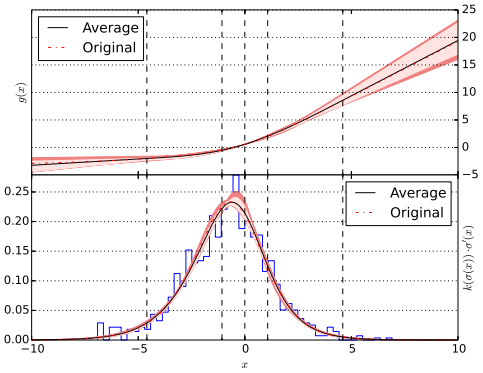


(c) $\tau = 0.8$

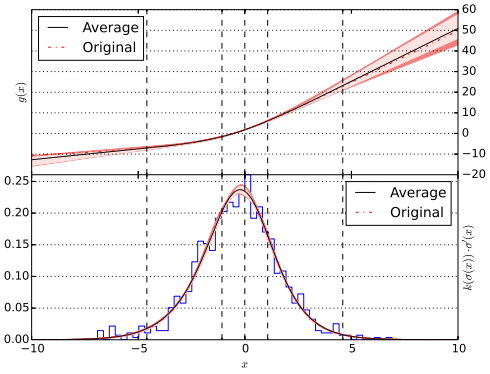
Figure 5.2: Frank copula g-generator and Kendall density



(a) $\tau = 0.2$

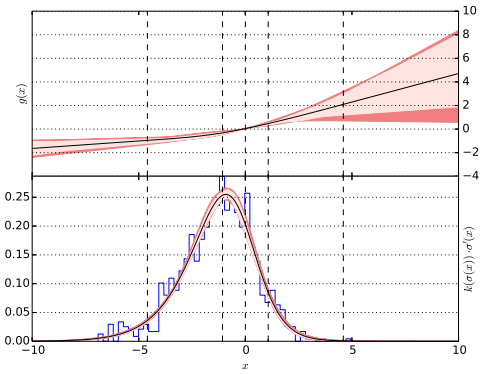


(b) $\tau = 0.5$

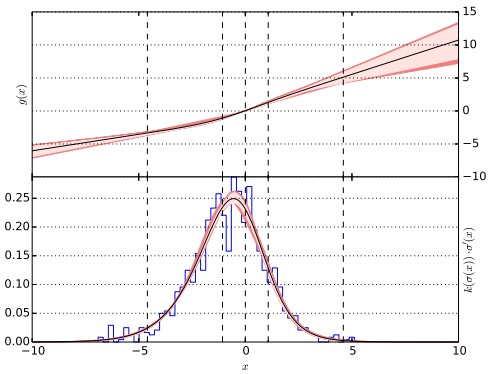


(c) $\tau = 0.8$

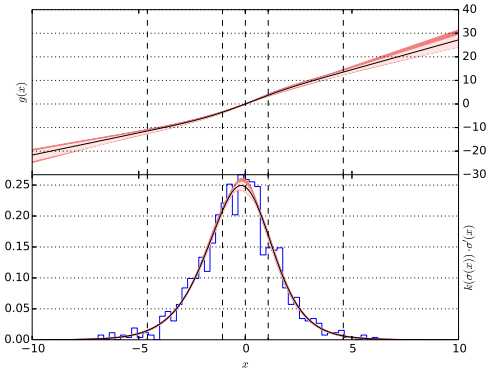
Figure 5.3: Gumbel copula g-generator and Kendall density



(a) $\tau = 0.2$

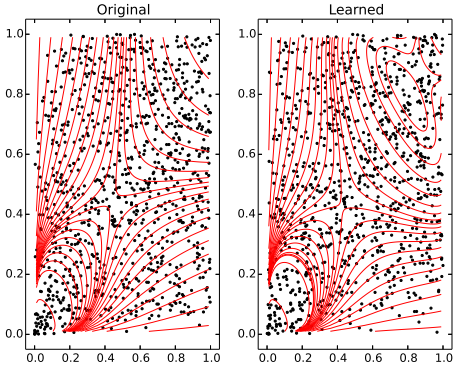


(b) $\tau = 0.5$

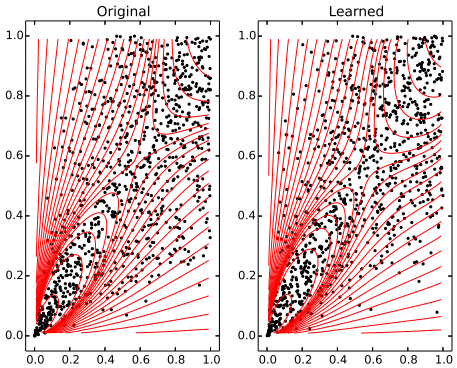


(c) $\tau = 0.8$

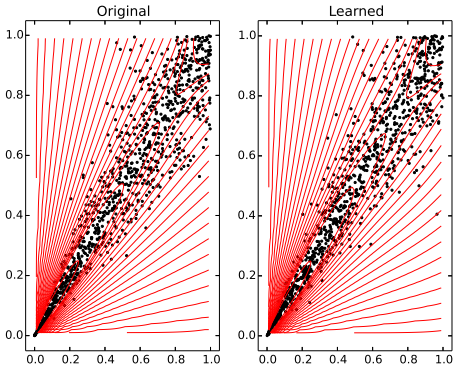
Figure 5.4: Gaussian copula g-generator and Kendall density



(a) $\tau = 0.2$

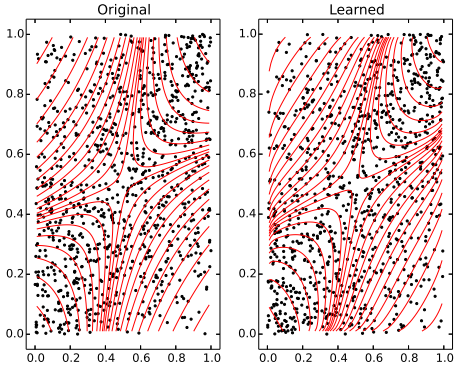


(b) $\tau = 0.5$

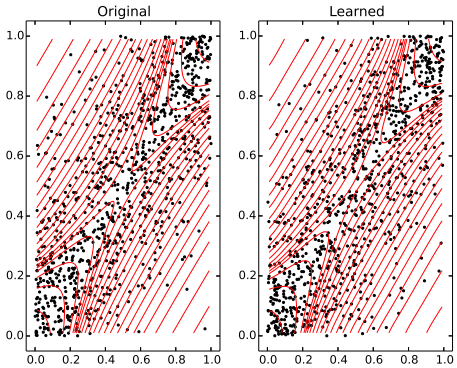


(c) $\tau = 0.8$

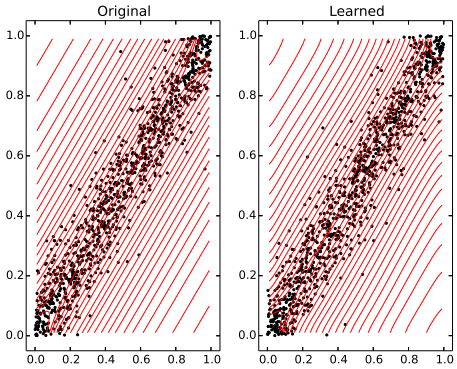
Figure 5.5: Clayton copula samples



(a) $\tau = 0.2$

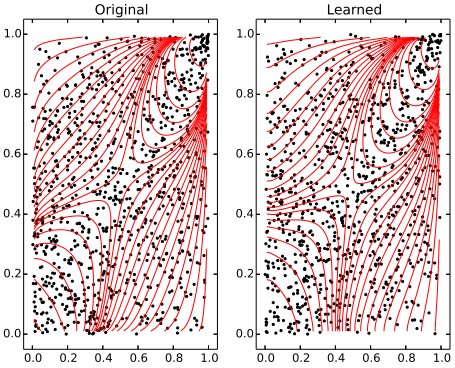


(b) $\tau = 0.5$

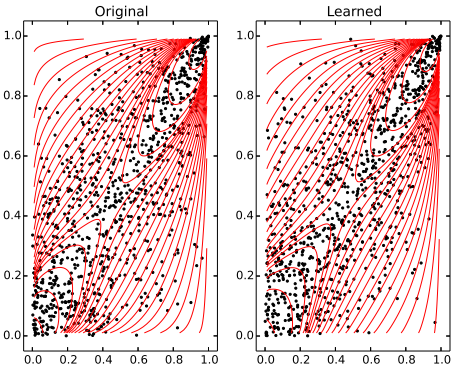


(c) $\tau = 0.8$

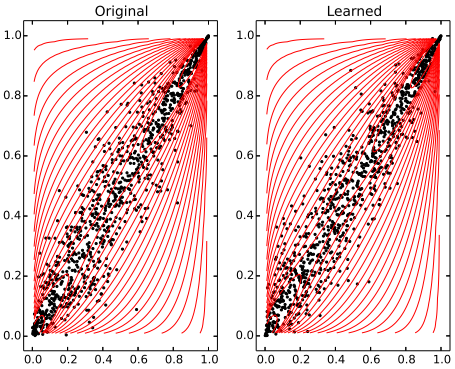
Figure 5.6: Frank copula samples



(a) $\tau = 0.2$

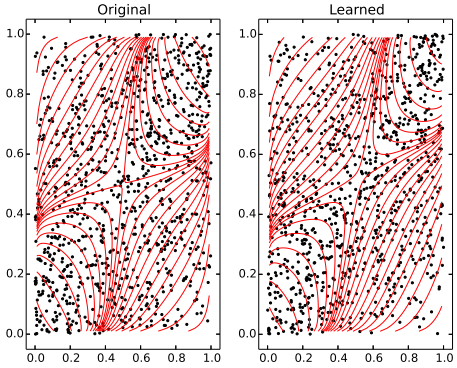


(b) $\tau = 0.5$

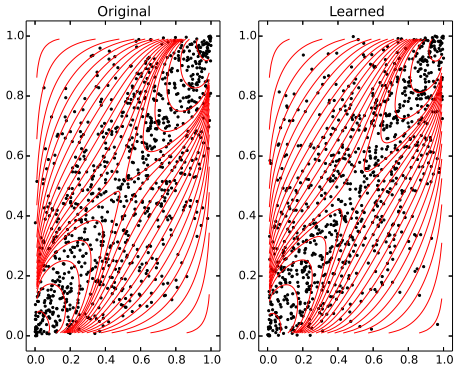


(c) $\tau = 0.8$

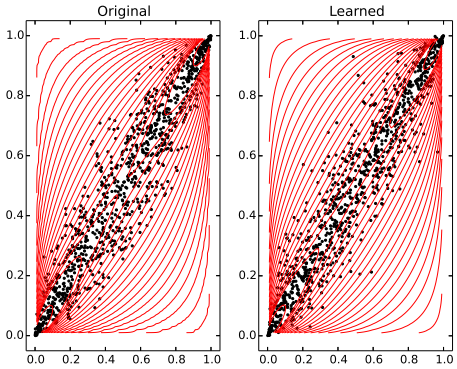
Figure 5.7: Gumbel copula samples



(a) $\tau = 0.2$



(b) $\tau = 0.5$



(c) $\tau = 0.8$

Figure 5.8: Gaussian copula samples

5.2 Trivariate copula simulations

In this section we will present some simulation examples that illustrate the feasibility and the strengths and weaknesses of our proposal for conditional copula modelling.[4]

5.2.1 Simulation settings and procedure

Our new technique for conditional copula modelling can be used to estimate 3-variate distributions via *vine* constructions, as explained in Section 3.2. We will test our method with synthetic data arising from 3-vines (Definition 2.6.1) whose bivariate margins C_{12} and C_{23} are set to the independence copula $C^{\perp}$. We will assume a perfect fit of those bivariate margins and focus on the estimation of the conditional copula family $\{C_{13|2}(*,*|w)\}_{w\in[0,1]}$. Examples from both Archimedean (namely Clayton, Frank and Gumbel) and non-Archimedean (Gaussian) families will be considered for the conditional copula family.

We know from the previous section that all the above families can be indexed in terms of Kendall's tau. In the conditional scenario, there is a function $\tau : [0,1] \to [-1,1]$ mapping each value w of the covariate to a Kendall's tau value that corresponds to a particular instance from a copula family. This way, the conditional copula family can be formulated as

$$C_{13|2}(u,v|w) = C_{\tau(w)}(u,v) \ , \tag{5.3}$$

where $\{C_\tau\}_{\tau\in[-1,1]}$ is any of the aforementioned parametric copula families.

We will take (Lambert, 2014), reviewed in Section 3.2.3, as a reference and use the same τ function he employs:

$$\tau(w) = 0.5 + 0.3\sin(1.6\pi w^{3/2}) \ . \tag{5.4}$$

Given a conditional copula family, we shall simulate $S = 10$ datasets $\mathcal{D}_i = \left\{\left(U_k^{(i)}, V_k^{(i)}, W_k^{(i)}\right)\right\}_{k=1}^{N}$, for $i = 1, 2, \ldots, S$, of $N = 2000$ trivariate points from that 3-vine, where the third component W represents the covariate. From each $\mathcal{D}_i$ we estimate several tensor product spline g-generators $\tilde{g}_i^{\beta}$ depending on a smoothing penalty algorithm parameter $\beta \geq 0$. Seven knots were employed for both the x and w variables of the spline generator, i.e., $n = 7$ and $m = 9$, for a total of $(n-1)m = 54$ parameters. Strictness was enforced in all Archimedean copulas (5.3).

Several kinds of graphical proofs are provided. One type aims at evaluating how successfully the estimation process fitted the original Archimedean g-generators $\{\tilde{g}_i^{\beta}(*,w)\}_{w\in[0,1]}$. This is the case of Figures 5.9, 5.10, 5.11, where the average $\bar{g}_\beta$ of the $\tilde{g}_i^{\beta}$'s is compared to the original g-generators $\{g(*,w)\}_{w\in[0,1]}$. Adding a function that only depends on w to the bivariate g-generators makes no difference. Therefore, we adjust the average $\bar{g}_\beta$ so that $\bar{g}_\beta(0,w) = g(0,w)$, defining

$$\bar{g}_\beta(x,w) = \frac{1}{S}\sum_{i=1}^{S}[\tilde{g}_i^{\beta}(x,w) - \tilde{g}_i^{\beta}(0,w) + g(0,w)] \ .$$

This way, results visualization is enhanced.

[4] See Section 4.2.

Besides the Archimedean generators, we will also check how well some dependence measures are fitted, namely Kendall's tau and both tail dependence indices. Let $\tilde{\tau}_i^{\beta}(w)$, $(\tilde{\lambda}_L)_i^{\beta}(w)$, $(\tilde{\lambda}_U)_i^{\beta}(w)$ be the Kendall's tau, lower tail index and upper tail index, respectively, of the i-th estimate of $C_{\tau(w)}$ and consider the averages

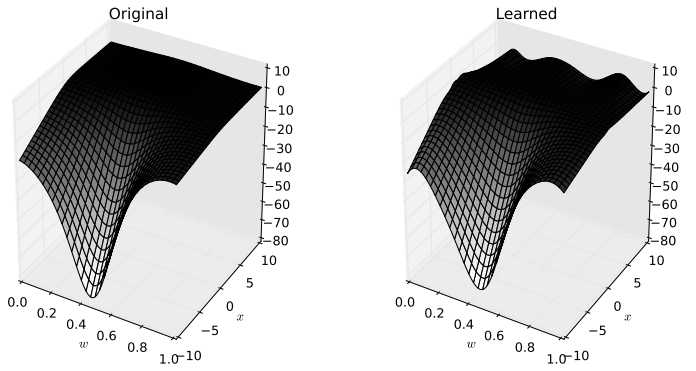
$$\bar{\tau}_{\beta}(w) = \frac{1}{S}\sum_{i=1}^{S} \tilde{\tau}_i^{\beta}(w) \ , \tag{5.5}$$

$$\bar{\lambda}_L^{\beta}(w) = \frac{1}{S}\sum_{i=1}^{S} (\tilde{\lambda}_L)_i^{\beta}(w) \ , \tag{5.6}$$

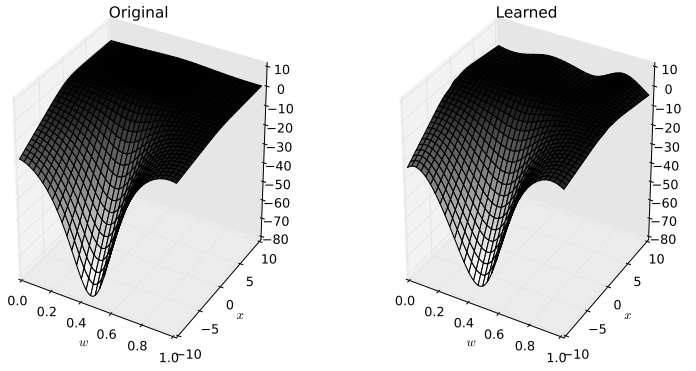
and

$$\bar{\lambda}_U^{\beta}(w) = \frac{1}{S}\sum_{i=1}^{S} (\tilde{\lambda}_U)_i^{\beta}(w) \ . \tag{5.7}$$

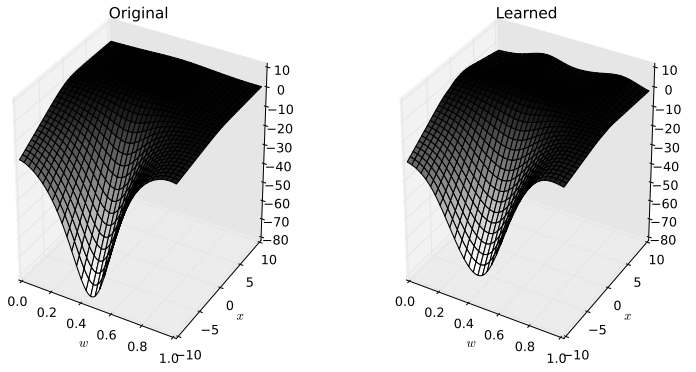
Figures 5.12 to 5.23 show the averages (5.5), (5.6) and (5.7), as well as the maximum, minimum and the 0.05 and 0.95 sample quantiles of $\tilde{\tau}_i^{\beta}(w)$, $(\tilde{\lambda}_L)_i^{\beta}(w)$ and $(\tilde{\lambda}_U)_i^{\beta}(w)$, much like in Figure 5.1.



(a) $\beta = 0$

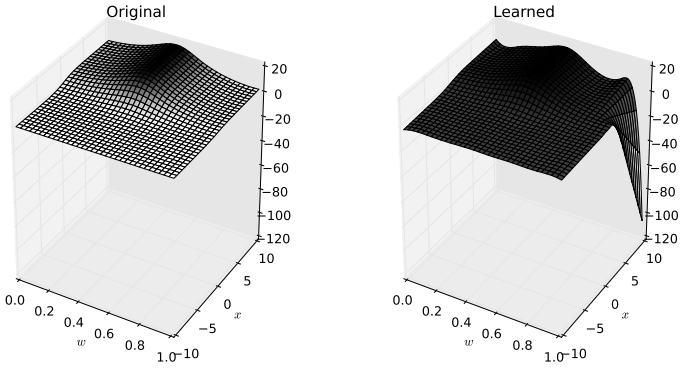


(b) $\beta = 10^{-9}$

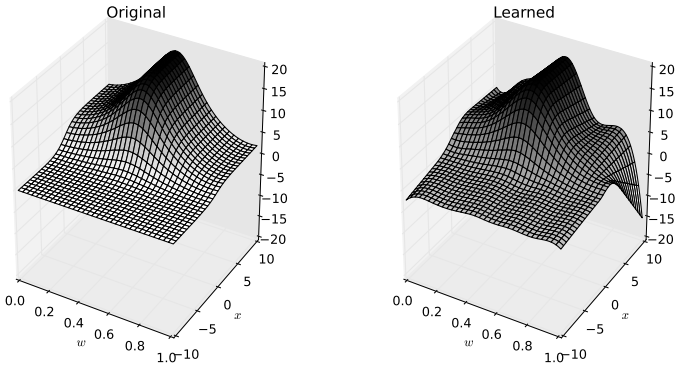


(c) $\beta = 10^{-8}$

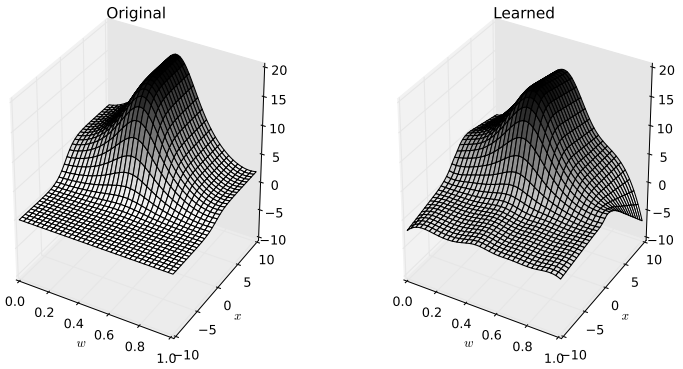
Figure 5.9: Clayton conditional family g-generators



(a) $\beta = 0$

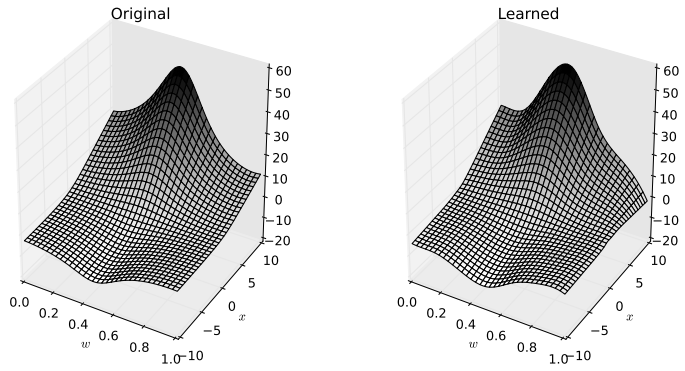


(b) $\beta = 10^{-9}$

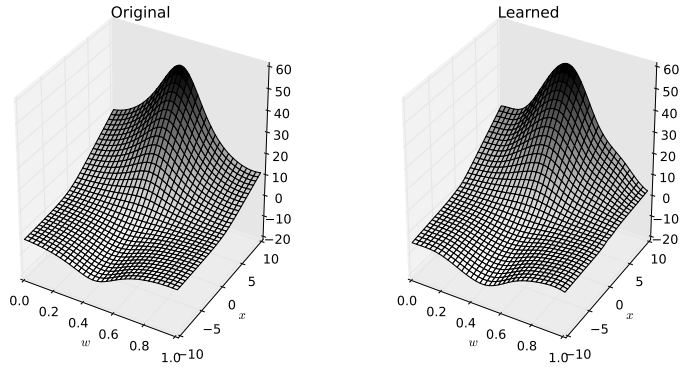


(c) $\beta = 10^{-8}$

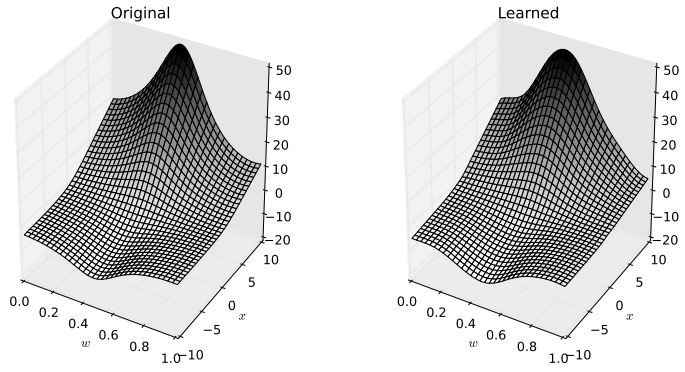
Figure 5.10: Frank conditional family g-generators



(a) $\beta = 0$

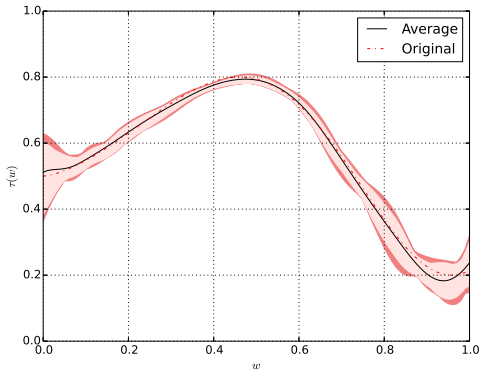


(b) $\beta = 10^{-9}$

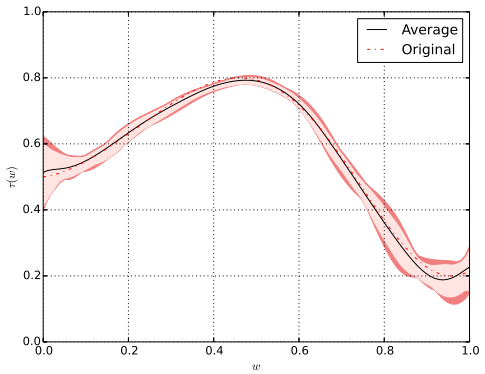


(c) $\beta = 10^{-8}$

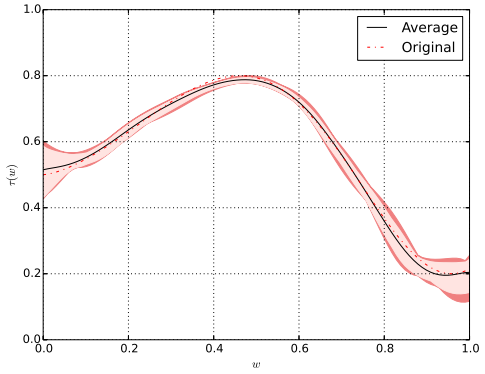
Figure 5.11: Gumbel conditional family g-generators



(a) $\beta = 0$

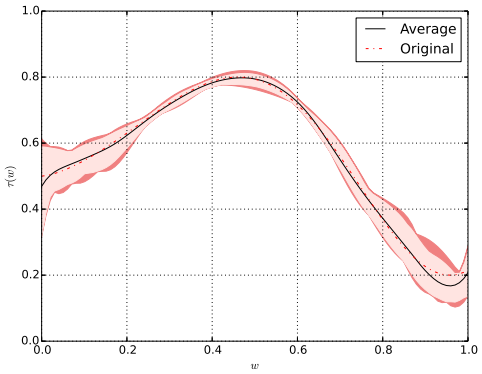


(b) $\beta = 10^{-9}$

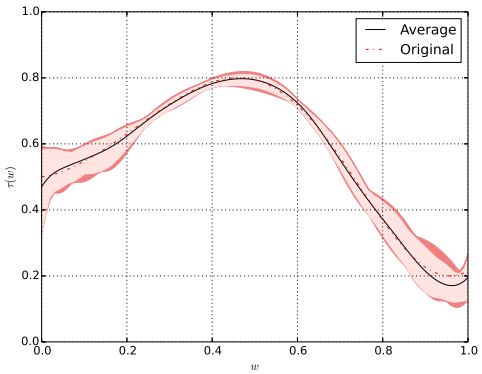


(c) $\beta = 10^{-8}$

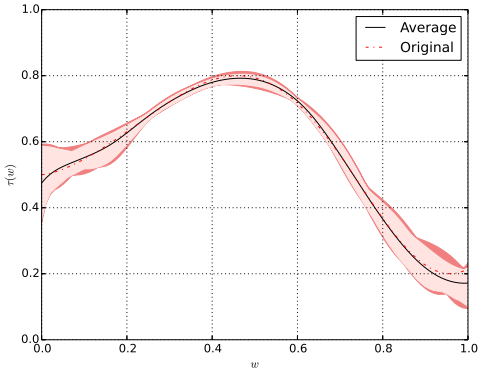
Figure 5.12: Clayton conditional family Kendall's tau



(a) $\beta = 0$

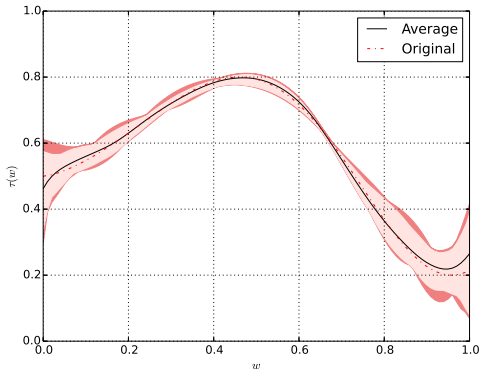


(b) $\beta = 10^{-9}$

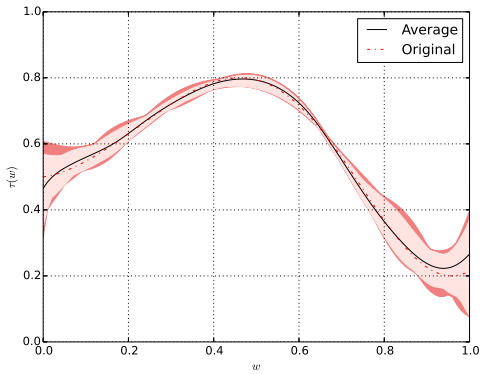


(c) $\beta = 10^{-8}$

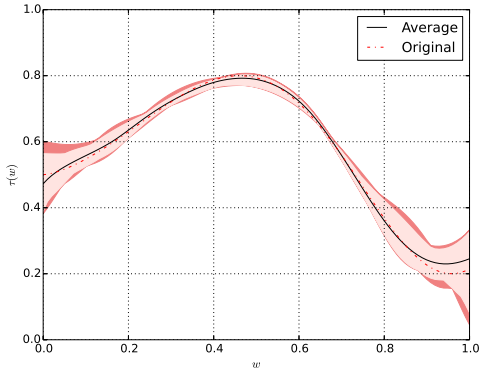
Figure 5.13: Frank conditional family Kendall's tau



(a) $\beta = 0$

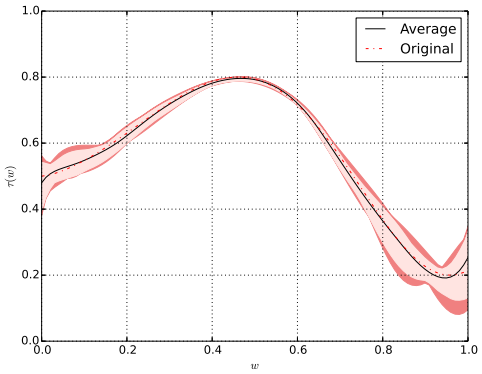


(b) $\beta = 10^{-9}$

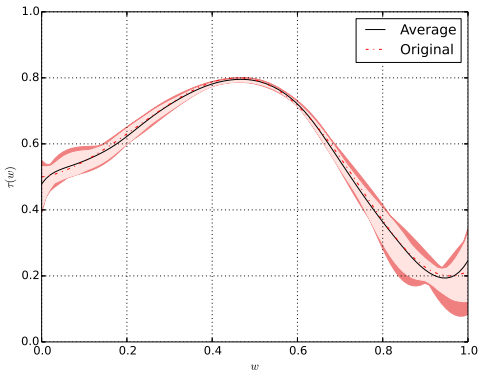


(c) $\beta = 10^{-8}$

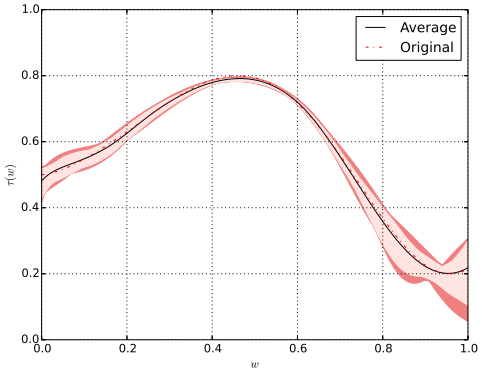
Figure 5.14: Gumbel conditional family Kendall's tau



(a) $\beta = 0$

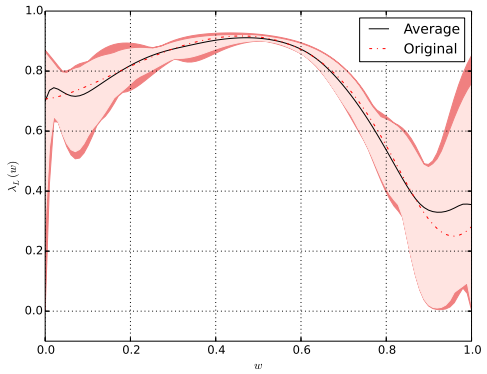


(b) $\beta = 10^{-9}$

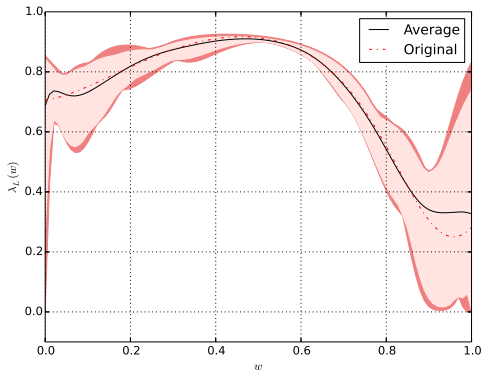


(c) $\beta = 10^{-8}$

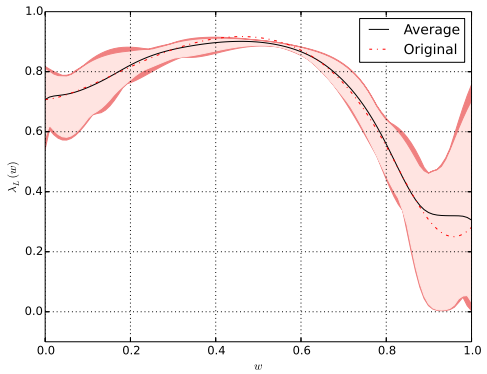
Figure 5.15: Gaussian conditional family Kendall's tau



(a) $\beta = 0$

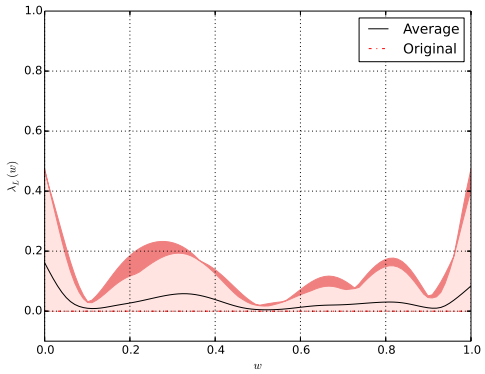


(b) $\beta = 10^{-9}$

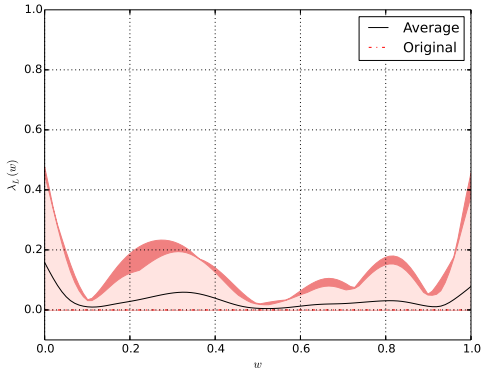


(c) $\beta = 10^{-8}$

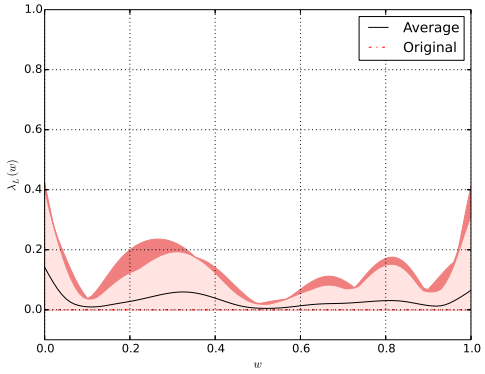
Figure 5.16: Clayton conditional family lower tail index



(a) $\beta = 0$

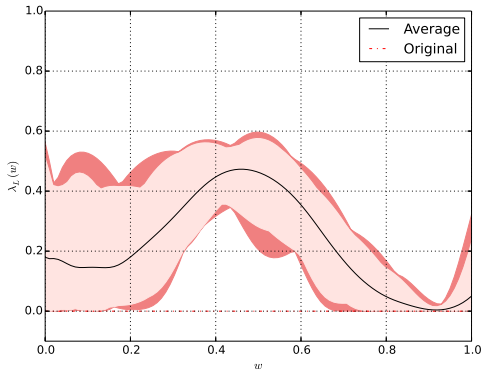


(b) $\beta = 10^{-9}$

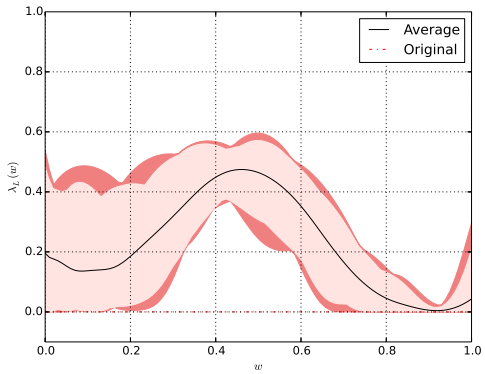


(c) $\beta = 10^{-8}$

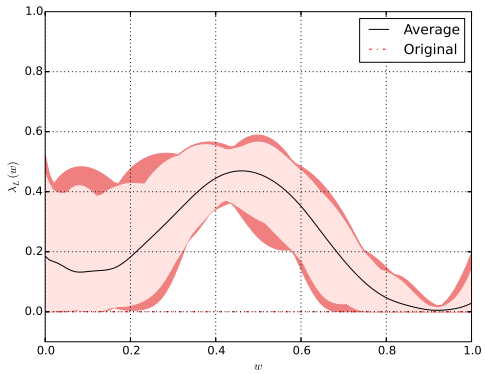
Figure 5.17: Frank conditional family lower tail index



(a) $\beta = 0$

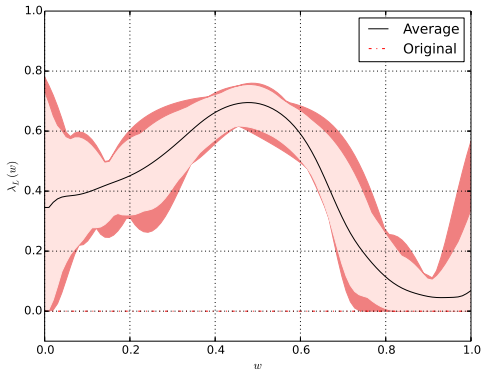


(b) $\beta = 10^{-9}$

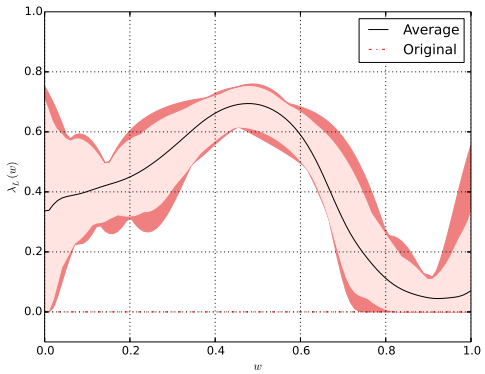


(c) $\beta = 10^{-8}$

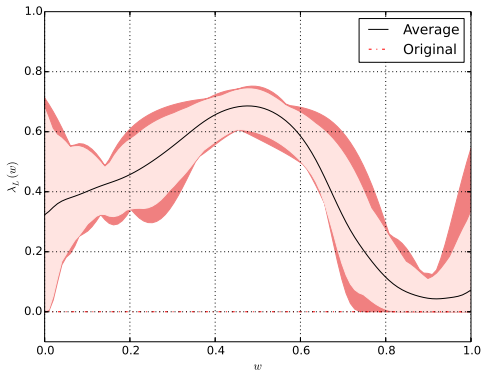
Figure 5.18: Gumbel conditional family lower tail index



(a) $\beta = 0$

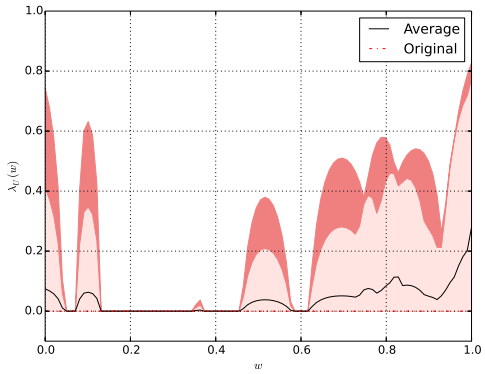


(b) $\beta = 10^{-9}$

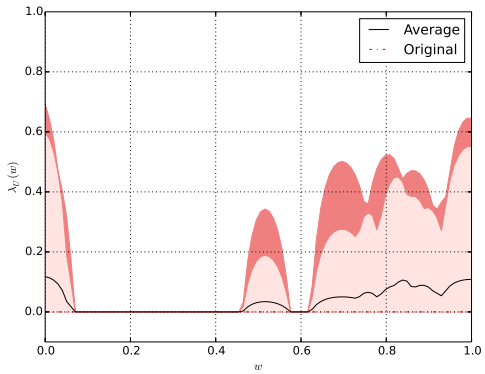


(c) $\beta = 10^{-8}$

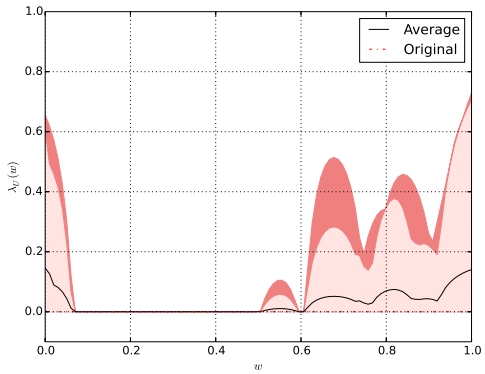
Figure 5.19: Gaussian conditional family lower tail index



(a) $\beta = 0$

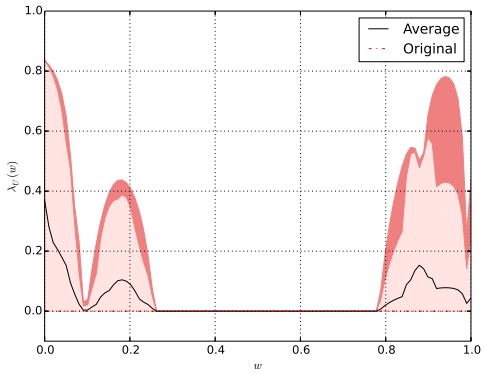


(b) $\beta = 10^{-9}$

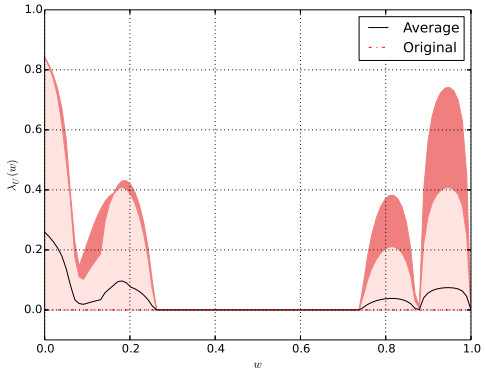


(c) $\beta = 10^{-8}$

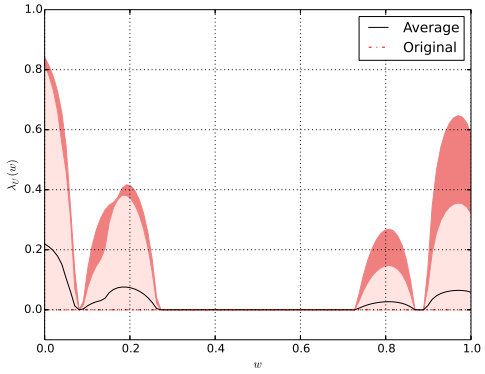
Figure 5.20: Clayton conditional family upper tail index



(a) $\beta = 0$

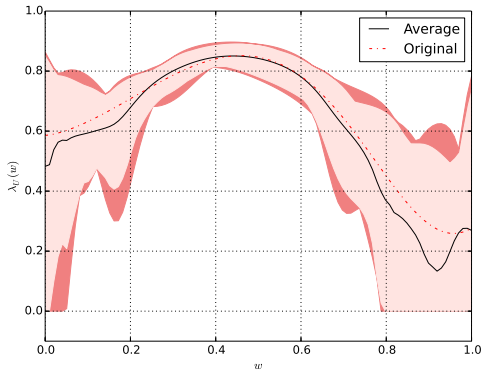


(b) $\beta = 10^{-9}$

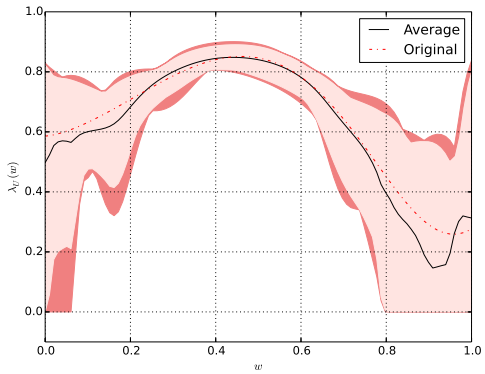


(c) $\beta = 10^{-8}$

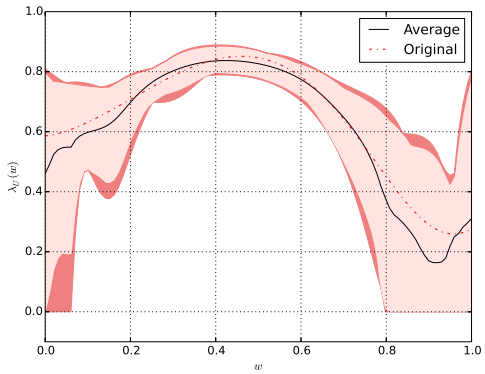
Figure 5.21: Frank conditional family upper tail index



(a) $\beta = 0$

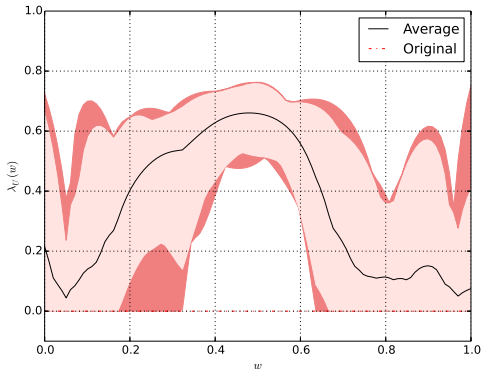


(b) $\beta = 10^{-9}$

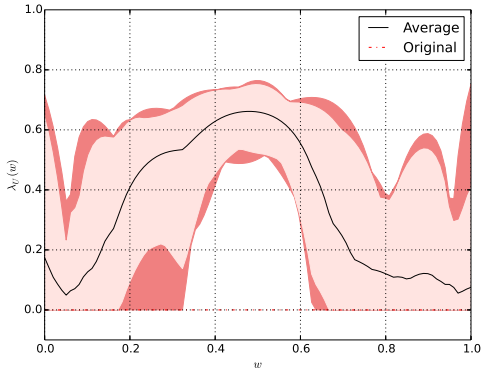


(c) $\beta = 10^{-8}$

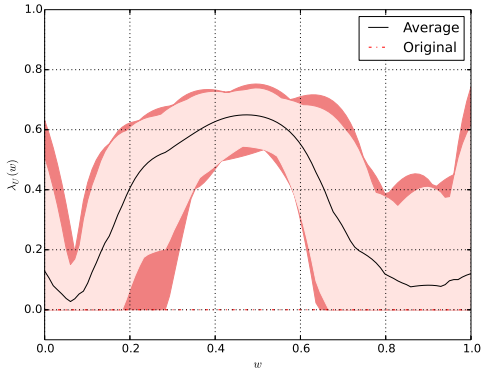
Figure 5.22: Gumbel conditional family upper tail index



(a) $\beta = 0$



(b) $\beta = 10^{-9}$



(c) $\beta = 10^{-8}$

Figure 5.23: Gaussian conditional family upper tail index

5.2.2 Comments

Figures 5.9, 5.10, 5.11 demonstrate the feasibility of our proposal. Despite the *curse of dimensionality* (Lopez-Paz et al., 2013), the overall appearance of the generators is reasonably well captured, except for the tails, if anything. As expected, due to the shortage of data, overfitting becomes apparent in the form of a curvature excess when $\beta = 0$, specially in Figures 5.9, 5.10. We can see that increasing the curvature penalty β the function gains some extra smoothness.

Figures 5.12, 5.14, 5.13, 5.15 show that the τ function (5.4) is, on average, fairly well reconstructed. Leaving out some procedural details, those plots are qualitatively similar to the ones presented in (Lambert, 2014). Both of us use $N = 2000$ sample points in the simulation study, but our respective approaches are essentially different. Lambert needs to optimize fewer parameters, despite employing a large number of knots for both variables, thanks to several additive parameter strategies. By contrast, at this stage of our research, we focus on a less sophisticated parameter configuration. On the other hand, our respective smoothing strategies for the simulation study are also different. Lambert uses an automatic Bayesian smoothing, while we opt for a penalized average log-likelihood, which allows us to graphically assess the effect of varying the smoothness degree.

(Lambert, 2014) does not report results on the estimation of tail dependence, so we cannot make any comparison at this respect. Our results are moderately positive, bearing in mind that modelling tail dependence is always a difficult task, specially in a 3-variate setting like ours. By far, the *worse* results we obtained were those of the Gaussian family, which is *not* Archimedean. Figures 5.19, 5.23 show that our method detects strong upper and lower tail dependencies when there is actually asymptotic independence in both tails. The Frank family has no tail dependence either and our method, again, erroneously captures tail dependence (Figures 5.17, 5.21), but this time by a narrow margin. Finally, the results for the Clayton (Figures 5.16, 5.20) and Gumbel (Figures 5.18, 5.22) families are fairly good. In particular, when there exists lower tail dependence (Figure 5.16) or upper tail dependence (Figure 5.22), it is fairly well estimated.

Chapter 6

Conclusions and future work

> – No ha dicho cuándo volverá.
> Hardin contestó:
> – Lo sé...; ¡pero espero que no vuelva hasta que usted y yo estemos segura y cómodamente muertos!
>
> *Fundación*
> Isaac Asimov

Throughout this work we have presented both a broad overview of multivariate copula modelling techniques and an in-depth new proposal on this matter. Chapter 2 and Chapter 3 gather some basic copula theory principles and some selected pieces from the state of the art, respectively, that provide the context for our research, the bulk of which is concentrated in Chapter 4 and Chapter 5.

The first part of Chapter 4 is devoted to expanding on (Hernández-Lobato and Suárez, 2011), adding some proofs and computational and implementation details (expanded in Appendix D) that Hernández-Lobato and Suárez did not originally include in their paper. We also propose some improvements regarding spline design (namely, spline knot locations and natural spline parametrization) and remark some new facts about intermediate tail dependence in the context of (Hernández-Lobato and Suárez, 2011).

The second part of Chapter 4 addresses a new semiparametric technique proposal that extends (Hernández-Lobato and Suárez, 2011) in a similar way to (Lambert, 2014). We construct a tensor product spline to fit a two-dimensional latent function, while specifying all relevant parameter constraints and some smoothing procedures based on a different roughness measure to the one used by Lambert.

The first part of Chapter 5 complements (Hernández-Lobato and Suárez, 2011) with newer simulations, while the second one illustrates, using synthetic data, the feasibility and effectiveness of our new proposal, obtaining qualitatively similar results to those in (Lambert, 2014).

From a theoretical standpoint, our proposal would benefit from an extensive and rigorous study of its convergence properties, which was out of the scope of

this work. From a more practical perspective, it would be desirable to test its flexibility and robustness on real data, as compared with methods like (Lambert, 2014). In order to make a fair comparison, we would need to adapt our model design and implementation so as to make it more efficient. By now, (Lambert, 2014) reports simulations using $S = 500$ samples, whereas we can reasonably handle up to $S = 10$ samples. Once matched the efficiency of (Lambert, 2014), we may extend our technique to more than one covariate, following perhaps Lambert's additive parameters approach.

Appendix A

Splines

This appendix is intended to summarize the main results about splines that we will extensively use throughout this work. Even though the basic concepts presented here constitute no new material, some notation, implementation and motivational details are due. See (Lyche and Morken, 2002) for a *friendly* (but complete) introduction and (Schumaker, 2007) for further insights.

A.1 Introduction

Definition A.1.1 (Spline). Given an interval $[a, b]$, with $a < b$, a function $f : [a, b] \to \mathbb{R}$ is called a spline of degree $d \geq 0$, represented by $f \in \mathbb{S}^d(\Delta)$, if it is left-continuous at b and there exists a partition $\Delta = \{\xi_i\}_{i=1}^n$ of $[a, b]$, with $a = \xi_1 < \xi_2 < \ldots < \xi_n = b$, such that the restriction of f to $[\xi_i, \xi_{i+1})$ is a polynomial of degree d, for $i = 1, 2, \ldots, n-1$.

From now on, we will use $\mathbb{S}_r^d(\Delta)$ as shorthand for $\mathbb{S}^d(\Delta) \cap \mathcal{C}^r(a, b)$, that is, the set of all splines with r-th continuous derivative.[1] Obviously, since $f \in \mathbb{S}^d(\Delta)$ satisfies $f_{|[\xi_i, \xi_{i+1})} \in \mathcal{C}^\infty(\xi_i, \xi_{i+1})$, the only troublesome points to ensure continuity are ξ_i, for $i = 2, 3, \ldots, n-1$. Moreover, notice that $\mathbb{S}_r^d(\Delta) = \mathcal{P}^d([a, b])$ if and only if $r \geq d$, so we will always assume $r \leq d$, with $r = d$ as a trivial limiting case.

The last definition simply formalizes the concept of piecewise polynomials. Other formalizations are equally valid. For instance, taking the interval in Definition $A.1.1$ to be left-open, instead of right-open, would also make sense for our purposes. Note that this is equivalent to asking for left continuity in place of right continuity, whenever full continuity cannot be assured. It must be stated, however, that a consistent selection of right or left continuity across all intervals eases spline representation, as we shall soon see. The inclusion of the endpoint b in the domain of a spline, along with the left-continuity assumption, though arbitrary, respond to practical considerations and have no effect on the algebraic structure of this set.

Splines play an important role in function approximation. Given a set of $n+1$ data points $\{(x_i, y_i)\}_{i=0}^n$ arising from a certain function sample, remember that there exists a unique polynomial of degree n that interpolates them. Generally

[1] Existence of the 0-th derivative means function continuity, by convention.

speaking, provided that the x-axis points are carefully chosen, interpolation error decreases as n increases, which in turn hampers computations.[2] Several spline solutions to the interpolation problem can also be found, with varying regularity at joints. Again, interpolation error decreases as partition Δ is finer, but, as opposed to global polynomial interpolation, local polynomial degrees remain fixed and so computational costs.[3]

Apart from avoiding *Runge's phenomenon* and speeding up computations, splines allow for greater flexibility when representing functions, while preserving simplicity.

Proposition A.1.1. *Let Δ be an n-point partition as in Definition A.1.1. We have that $\mathbb{S}_r^d(\Delta)$ is a vector space and*

$$\begin{aligned}\dim \mathbb{S}_r^d(\Delta) &= (n-1)d - (n-2)r + 1 \\ &= (d-r)n + 2r - d + 1 \ .\end{aligned}$$

Proof. First of all, it is easy to check that $\mathbb{S}^d(\Delta)$ is a vector space and $\mathbb{S}_r^d(\Delta)$ has closure under function addition and scalar multiplication: $\mathbb{S}_r^d(\Delta)$ is a vector subspace of $\mathbb{S}^d(\Delta)$. Secondly, $\mathbb{S}^d(\Delta)$ is isomorphic to the Cartesian product vector space of $\mathcal{P}^d(\mathbb{R})$ (real polynomials of degree d) with itself $n-1$ times (abbreviated $\left(\mathcal{P}^d(\mathbb{R})\right)^{n-1}$) through the linear map $\Pi : f \mapsto (p_1, p_2, \ldots, p_{n-1})$, where p_i is such that $f_{|[\xi_i,\xi_{i+1})}(x) = p_i(x) \cdot \mathbb{1}_{|[\xi_i,\xi_{i+1})}(x)$. Hence, $\dim \mathbb{S}^d(\Delta) = (n-1) \times (d+1)$.

Now, consider the linear function L that maps $f \in \mathbb{S}^d(\Delta)$ to the $(n-2) \times (r+1)$ real matrix

$$\begin{bmatrix} p_1(\xi_2) - p_2(\xi_2) & p_2(\xi_3) - p_3(\xi_3) & \cdots & p_{n-2}(\xi_{n-1}) - p_{n-1}(\xi_{n-1}) \\ p_1'(\xi_2) - p_2'(\xi_2) & p_2'(\xi_3) - p_3'(\xi_3) & \cdots & p_{n-2}'(\xi_{n-1}) - p_{n-1}'(\xi_{n-1}) \\ \vdots & \vdots & & \vdots \\ p_1^{r)}(\xi_2) - p_2^{r)}(\xi_2) & p_2^{r)}(\xi_3) - p_3^{r)}(\xi_3) & \cdots & p_{n-2}^{r)}(\xi_{n-1}) - p_{n-1}^{r)}(\xi_{n-1}) \end{bmatrix} ,$$

where the p_i's are defined as before. From basic Linear Algebra, we know that $\dim \mathbb{S}^d(\Delta) = \dim \mathrm{Ker}(L) + \dim \mathrm{Im}(L)$. Of course, $\mathrm{Ker}(L) = \mathbb{S}_r^d(\Delta)$, so it suffices to verify that $\dim \mathrm{Im}(L) = (n-2) \times (r+1)$, that is to say L is a surjective function.

We will show that L can be broken down into surjective linear maps as $L = \Xi \circ D^r \circ \Lambda \circ \Pi$, so it is also onto. Firstly, let

$$\begin{array}{rccc} \Lambda : & \left(\mathcal{P}^d(\mathbb{R})\right)^{n-1} & \longrightarrow & \left(\mathcal{P}^d(\mathbb{R})\right)^{n-2} \\ & (p_1, p_2, \ldots, p_{n-1}) & \mapsto & (p_1 - p_2, p_2 - p_3, \ldots, p_{n-2} - p_{n-1}) \end{array} .$$

Any basis of polynomials of degree d induces a basis both in $\left(\mathcal{P}^d(\mathbb{R})\right)^{n-1}$ and $\left(\mathcal{P}^d(\mathbb{R})\right)^{n-2}$. The matrix of Λ with respect to those basis is always $A \otimes I_{d+1}$,

[2] See *Runge's phenomenon* (Schumaker, 2007).

[3] Evaluating a spline at a point x requires finding $i \in \{1, 2, \ldots, n-1\}$ such that $x \in [\xi_i, \xi_{i+1})$. A sensible single-thread implementation would perform a *binary search* over Δ (an ordered sequence), with a relatively cheap $\mathcal{O}(\log n)$ order of comparison operations, plus a fixed number of additions and multiplications. By contrast, evaluating an n-degree polynomial requires, at best, $\mathcal{O}(n)$ additions and multiplications.

where

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 & -1 \end{bmatrix} \in \mathcal{M}_{(n-2)\times(n-1)}(\mathbb{R}) \ ,$$

so $\dim \mathrm{Im}(\Lambda) = \mathrm{rank}(A \otimes I_{d+1}) = \mathrm{rank}(A) \times \mathrm{rank}(I_{d+1}) = (n-2) \times (d+1)$, thus Λ is onto.

Given $x \in \mathbb{R}$, it easy to check that

$$\begin{array}{rccl} D^r_x : & \mathcal{P}^d(\mathbb{R}) & \longrightarrow & \mathbb{R}^{r+1} \\ & p & \mapsto & (p(x), p'(x), p''(x), \dots, p^{r)}(x)) \end{array}$$

is an onto linear map. Then, clearly,

$$\begin{array}{rccl} D^r : & \left(\mathcal{P}^d(\mathbb{R})\right)^{n-2} & \longrightarrow & \left(\mathbb{R}^{r+1}\right)^{n-2} \\ & p & \mapsto & \left(D^r_{\xi_2}(p), D^r_{\xi_3}(p), \dots, D^r_{\xi_{n-1}}(p)\right) \end{array}$$

is, too. Finally, let Ξ be the isomorphism $\left(\mathbb{R}^{r+1}\right)^{n-2} \cong \mathcal{M}_{(n-2)\times(r+1)}(\mathbb{R})$ transforming tuple vector components into matrix columns. □

From last proposition we see that the more regularity r we demand, the lower the space dimension is. Namely, for the cubic case ($d = 3$), we have

$$\dim \mathbb{S}^3_1(\Delta) = 2n \tag{A.1}$$

and

$$\dim \mathbb{S}^3_2(\Delta) = n + 2 \ . \tag{A.2}$$

Case (A.1) corresponds to *(Cubic) Hermite interpolation*, that is, constructing a cubic spline f by specifying $f(\xi_i)$ and $f'(\xi_i)$, for $i = 1, 2, \dots, n$. On the other hand, (A.2) tells us that, even if we fixed values $f(\xi_i)$, for $i = 1, 2, \dots, n$, the resulting vector space of cubic splines f would have dimension 2. Adding the following two linear boundary conditions

$$f''(a^+) = f''(b^-) = 0 \ , \tag{A.3}$$

we get the subspace of *natural* (cubic) splines, $\mathbb{S}_{\mathcal{N}}(\Delta)$. Imposing instead, for some $\alpha, \beta \in \mathbb{R}$,

$$f'(a^+) = \alpha \text{ and } f'(b^-) = \beta \ , \tag{A.4}$$

we get the (affine) subspace of *complete* or *clamped* (cubic) splines, $\mathbb{S}_{\alpha,\beta}(\Delta)$.

A.2 A basis for splines: B-splines

Proposition *A*.1.1 showed that the set of splines with a given regularity has vector space structure. The proof implicitly specified (through isomorphism) a possible basis for this space. Now, we explicitly state that $f \in \mathbb{S}^d(\Delta)$ has the form

$$f(x) = \sum_{i=1}^{n-1} \sum_{j=0}^{d} \lambda^i_j x^j \mathbb{1}_{I_i}(x) = \sum_{i=1}^{n-1} p_i(x) \mathbb{1}_{I_i}(x) \ , \tag{A.5}$$

for some scalars λ_j^i, where $I_i = [\xi_i, \xi_{i+1})$, for all $i < n-1$, and $I_{n-1} = [\xi_{n-1}, \xi_n]$. Despite its simplicity, the basis $\{x \mapsto x^j \mathbb{1}_{I_i}(x) \ : \ 1 \leq i \leq n-1 \text{ and } 0 \leq j \leq d\}$ does not have *built-in* smoothness: we have to impose on the variables in (A.5) the constraints

$$p_{i-1}^{k)}(\xi_i) = p_i^{k)}(\xi_i) \ , \tag{A.6}$$

for $i = 2, 3, \ldots, n-1$ and $k = 0, 1, \ldots, r$, in order for f to be r times differentiable.

The next family of functions, known as *B-splines* or *basis splines*, will provide a more general solution for representing splines.

Definition A.2.1 (B-splines). Let $d \in \mathbb{Z}_+$ and let $\mathbf{t} = (t_1, t_2, \ldots, t_{m+d+1})$ be a real vector, for $m \in \mathbb{N}$, such that $t_i \leq t_j$ whenever $i < j$. For each $k \in \mathbb{Z}$ between 0 and d and for each $j \in \mathbb{N}$ between 1 and $m + d - k$, we recursively define the j-th B-spline of degree k and knots $\mathbf{t}$, $B_{j,k}^{\mathbf{t}} : \mathbb{R} \to \mathbb{R}$, as

$$B_{j,k}^{\mathbf{t}}(x) = \omega_{j,k}^{\mathbf{t}}(x) B_{j,k-1}^{\mathbf{t}}(x) + \bar{\omega}_{j+1,k}^{\mathbf{t}}(x) B_{j+1,k-1}^{\mathbf{t}}(x) \ , \tag{A.7}$$

where the complementary weights are defined as

$$\begin{aligned} \omega_{i,k}^{\mathbf{t}}(x) &= \begin{cases} \frac{x - t_i}{t_{i+k} - t_i}, & \text{if } t_i \leq x < t_{i+k} \\ 0, & \text{otherwise} \end{cases} \\ \bar{\omega}_{i,k}^{\mathbf{t}}(x) &= 1 - \omega_{i,k}^{\mathbf{t}}(x) \end{aligned} \tag{A.8}$$

and finally

$$B_{j,0}^{\mathbf{t}} = \mathbb{1}_{[t_j, t_{j+1})} \ . \tag{A.9}$$

Several equivalent possibilities arise when defining B-splines. Inductively, we can see that the range of $B_{j,k}^{\mathbf{t}}$ is completely determined by knots $t_j, t_{j+1}, \ldots, t_{j+k+1}$ and, in fact, $B_{j,k}^{\mathbf{t}}$ vanishes outside $[t_j, t_{j+k+1})$. In particular, if $t_j = t_{j+k+1}$, then $B_{j,k}^{\mathbf{t}}(x) = 0$, for all $x \in \mathbb{R}$.[4] This allows to rewrite (A.7) simply as

$$B_{j,k}^{\mathbf{t}}(x) = \frac{B_{j,k-1}^{\mathbf{t}}(x)}{t_{j+k} - t_j}(x - t_j) + \frac{B_{j+1,k-1}^{\mathbf{t}}(x)}{t_{j+k+1} - t_{j+1}}(t_{j+k+1} - x) \ ,$$

assuming the convention that $0/0 = 0$.

Observe that the definition of B-splines can easily be stated in terms of an infinite (alternatively, doubly infinite) (de Boor, 1972) non-decreasing real sequence $\mathbf{t} = \{t_i\}_{i \in \mathbb{N}}$ (alternatively, $\mathbf{t} = \{t_i\}_{i \in \mathbb{Z}}$) automatically defining $B_{j,k}^{\mathbf{t}}$ for all $k \in \mathbb{Z}_+$ and for all $j \in \mathbb{N}$ (alternatively, $j \in \mathbb{Z}$). Using a vector instead restricts both the degree and the number of B-splines for that degree. Notwithstanding, given a degree d, all results in this section require a finite number of functions $B_{j,d}^{\mathbf{t}}$, typically requiring $m > d$; one of the axioms of an *extended* knot vector, according to (Lyche and Morken, 2002). Moreover, we shall be mainly concerned with cubic B-splines ($d = 3$).

B-splines are indeed splines, as defined in Definition $A.1.1$, and the notion of degree in both cases also coincides. This can easily been demonstrated by induction on degree k. More precisely, we claim that the restriction of $B_{j,k}^{\mathbf{t}}$ to a

[4] Indeed, if $t_j = t_{j+1}$ in (A.9), then $[t_j, t_{j+1}) = \emptyset$ and $B_{j,0}^{\mathbf{t}}$ is identically zero.

certain partitioned interval is a spline of degree k. Namely, if $t_j < t_{j+k+1}$, then the partition we look for is $\{t_j, t_{j+1}, \ldots, t_{j+k+1}\}$ (over interval $[t_j, t_{j+k+1})$); if, on the contrary, $t_j = t_{j+k+1}$, then $B^{\mathbf{t}}_{j,k}$ turns out to be identically zero and thus any partitioned interval suffices.

Several other properties of B-splines can be inductively checked. The following, which can actually be derived as a corollary of a more general result, is one of them.

Proposition A.2.1. *Let d, m and $\mathbf{t}$ be as in Definition A.2.1 and assume $m > d$. We have*

$$\sum_{j=1}^{m} B^{\mathbf{t}}_{j,d}(x) = 1, \text{ for all } x \in [t_{d+1}, t_{m+1}) \ .$$

Proof. It easily follows by induction, bearing in mind that $B^{\mathbf{t}}_{1,d-1}$ and $B^{\mathbf{t}}_{m+1,d-1}$ vanish outside $[t_1, t_{d+1})$ and $[t_{m+1}, t_{m+d+1})$, respectively. □

From now on, we will omit the knot vector superscript $\mathbf{t}$, both in B-splines and in the weighting functions involved, when there is no place for confusion. Next, we introduce an easy result that will considerably simplify calculations with B-splines.

Proposition A.2.2. *Let d, m and $\mathbf{t}$ be as in Definition A.2.1 and let $z \in \mathbb{R}$. If there exists $1 \leq j \leq m$ such that $z = t_{j+1} = t_{j+2} = \ldots = t_{j+d} < t_{j+d+1}$, then $B^{\mathbf{t}}_{j,d}(z) = 1$ and $B^{\mathbf{t}}_{i,d}(z) = 0$ if $i \neq j$.*

Proof. Let us prove it by induction on degree d, starting with $d = 1$. Suppose that $z = t_{j+1} < t_{j+2}$. Then we have $B_{j+1,0}(z) = 1$ and $B_{i,0}(z) = 0$ for $i \neq j+1$. Therefore,

$$B_{j,1}(z) = \frac{t_{j+2} - z}{t_{j+2} - t_{j+1}} B_{j+1,0}(z) = 1 \ ,$$

no matter $t_j < t_{j+1}$ or not. To see $B_{i,1}(z) = 0$ whenever $i \neq j$, note that the latter builds upon $B_{i,0}$ and $B_{i+1,0}$, thus $i = j+1$ is needed for it to be greater than zero. But then,

$$\frac{z - t_i}{t_{i+1} - t_i} = \frac{t_{j+1} - t_{j+1}}{t_{j+2} - t_{j+1}} = 0 \ ,$$

so $B_{i,1}(z) = 0$, regardless.

Now, suppose the statement holds for $d-1$. Taking $l = j+1$ and $k = d-1$, we have $z = t_{l+1} = t_{l+2} = \ldots = t_{l+k} < t_{l+k+1}$ so, by inductive hypothesis, $B_{l,k}(z) = 1$ and $B_{l-1,k}(z) = 0$, hence

$$B_{j,d}(z) = \omega_{j,d}(z) B_{l-1,k}(z) + \bar{\omega}_{j+1,d}(z) B_{l,k}(z) = 1 \ .$$

Finally, for $B_{i,d}(z) = 1$ we need $i + 1 = l$, that is, $i = j$. □

Now it is time to address some differentiability properties of B-splines.

Proposition A.2.3. *Allowing the usual convention that $0/0 = 0$, the r-th right derivative[5] of a basis spline function of degree $d \geq 1$ satisfies the following*

[5] Of course, the only points where the right derivative of a spline might not be equal to its left derivative are the knots in $\mathbf{t}$.

recurrence relation:

$$D^r B_{j,d} = d \cdot \left(\frac{D^{r-1} B_{j,d-1}}{t_{j+d} - t_j} - \frac{D^{r-1} B_{j+1,d-1}}{t_{j+d+1} - t_{j+1}} \right) ,$$

where $D^0 f(x) = f(x^+)$.

Proposition A.2.4. *Let* d *and* $\mathbf{t}$ *be as in Definition A.2.1. Let* $z \in \mathbb{R}$ *such that* $z = t_{i_k}$ *for* $j \leq i_1 < i_2 < \ldots < i_n \leq j + d + 1$. *Then,* $B^{\mathbf{t}}_{j,d}$ *has continuous* r*-th derivative at* z *for* $r = 0, 1, \ldots, d - n$.

Definition *A*.2.2 and Theorem *A*.2.1 give us the keys to represent splines as linear combinations of B-splines, which can be efficiently and stably evaluated (de Boor, 1972).

Definition A.2.2. Let d, m and $\mathbf{t}$ be as in Definition *A*.2.1. We define the vector space of linear combinations of B-splines as

$$\mathbb{B}^d(\mathbf{t}) = \operatorname{span}\{B^{\mathbf{t}}_{1,d}, B^{\mathbf{t}}_{2,d}, \ldots, B^{\mathbf{t}}_{m,d}\} = \left\{ x \in \mathbb{R} \mapsto \sum_{i=1}^{m} \theta_i B^{\mathbf{t}}_{i,d}(x) : \theta_i \in \mathbb{R} \right\} .$$

Theorem A.2.1. *Let* Δ *be an* n*-point partition of* $[a, b]$ *as in Definition A*.1.1. *Consider the extended knot vector*

$$\mathbf{t} = (\mathbf{a};\ \overbrace{\xi_2, \ldots, \xi_2}^{d-r \text{ times}}, \overbrace{\xi_3, \ldots, \xi_3}^{d-r \text{ times}}, \ldots, \overbrace{\xi_{n-1}, \ldots, \xi_{n-1}}^{d-r \text{ times}};\ \mathbf{b}) , \tag{A.10}$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_{d+1}) \in \mathbb{R}^{d+1}$ *and* $\mathbf{b} = (b_1, b_2, \ldots, b_{d+1}) \in \mathbb{R}^{d+1}$ *satisfy* $a_1 \leq a_2 \leq \ldots \leq a_{d+1} \leq a$ *and* $b \leq b_1 \leq b_2 \leq \ldots \leq b_{d+1}$. *Splines and linear combinations of B-splines coincide when restricted to* $[a, b)$, *that is*

$$\mathbb{S}^d_r(\Delta)_{|[a,b)} = \mathbb{B}^d(\mathbf{t})_{|[a,b)} . \tag{A.11}$$

Proof. The key result for this proof, which we will not include in this annex, is the linear independence of functions $B^{\mathbf{t}}_{1,d}, B^{\mathbf{t}}_{2,d}, \ldots, B^{\mathbf{t}}_{m,d}$, given an *extended* knot vector as (A.10), when they are all restricted to $[t_{d+1}, t_{m+1})$, with $m > d$ (Lyche and Schumaker, 1975, Schumaker, 2007, Lyche and Morken, 2002). Since $t_{d+1} = a_{d+1} \leq a$ and $b \leq b_1 = t_{m+1}$, they are also linearly independent when restricted to $[a, b)$. Additionally, for Proposition *A*.2.4, $\mathbb{B}^d(\mathbf{t})_{|[a,b)} \subset \mathbb{S}^d_r(\Delta)_{|[a,b)}$. Finally, $m + d + 1 = |\mathbf{t}| = (n - 2) \times (d - r) + 2 \times (d + 1)$, so $\dim \mathbb{B}^d(\mathbf{t})_{|[a,b)} = m = \dim \mathbb{S}^d_r(\Delta)_{|[a,b)}$, which ends the proof. □

The selection of the knots a_i and b_i, for $i = 1, 2, \ldots, d + 1$, provided that they satisfy the hypotheses given in Theorem *A*.2.1, has no effect on (A.11). For practical reasons, as we shall soon see, a convenient selection is $a_i = a$ and $b_i = b$, for all i, in which case $\mathbf{t}$ is called $(d + 1)$-*regular* (Lyche and Morken, 2002). Notice that we would much rather have

$$\mathbb{S}^d_r(\Delta) = \mathbb{B}^d(\mathbf{t})_{|[a,b]} , \tag{A.12}$$

than (A.11), so that splines (and their derivatives, according to Proposition *A*.2.3 could be evaluated as a linear combination of B-splines at endpoint

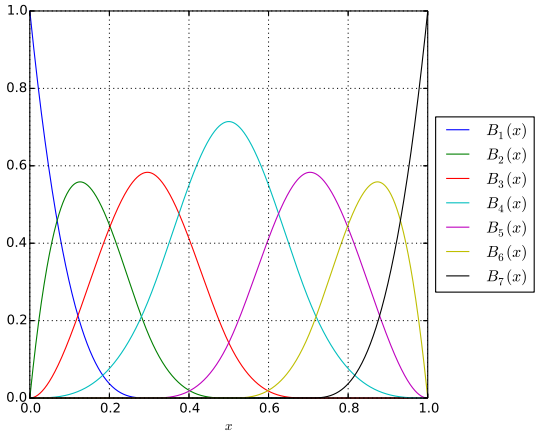


Figure A.1: B-splines for $\mathbb{S}_2^3(\Delta)$, where $\Delta = \{0, 0.3, 0.5, 0.7, 1.0\}$.

b too. This would be granted if $b < b_1$; with $(d+1)$-regularity, though, we need to redefine (A.9) for $j = m$:

$$B_{m,0}^{\mathbf{t}} = \mathbb{1}_{[t_m, b]} \ . \tag{A.13}$$

We shall assume this workaround from now on. Figure A.1 shows the B-splines for $\mathbb{S}_2^3(\Delta)$ according to Theorem *A*.2.1.

Proposition A.2.5. *Let d, m and r be the usual parameters for a smooth spline. Let $\mathbf{t}$ be as in (A.10), with the additional assumption that $a = a_i$ and $b = b_i$, for all $i = 1, 2, \ldots, d+1$, and the subsequent redefinition (A.13). Then $B_{m,k}^{\mathbf{t}}(b) = 1$ and $B_{j,k}^{\mathbf{t}}(b) = 0$ if $j \neq m$, for $k = 0, 1, \ldots, d$.*

Proof. It easily follows by induction on k. □

Many times one wants to extend a spline function out of its domain of definition in a sensitive manner. One way to do this is by linear extrapolation, defining:

$$\begin{aligned} f(x) &= f(a) + f'(a^+)(x - a) \\ f(x) &= f(b) + f'(b^-)(x - b) \end{aligned} \ , \tag{A.14}$$

for $x \leq a$ and $x \geq b$, respectively. This strategy results specially convenient when the natural boundary conditions (A.3) are met, yielding a legitimate $\mathcal{C}^2(\mathbb{R})$ function (see Figure A.3). The following result gives us simple expressions for the previous evaluations of f and its derivatives in terms of basis spline dimension, coordinates and knots.

Proposition A.2.6. *Let Δ and $f \in \mathbb{S}_r^d(\Delta)$, with coordinates $\theta_1, \theta_2, \ldots, \theta_m$, like in Theorem A.2.1. Suppose, for the sake of simplicity, maximum non-trivial regularity r, $r = d - 1$, so that every interior point $\xi_i \in \Delta$, $1 < i < n$, appears just once in the extended knot vector (A.10). Taking $\delta_i^a = a - \xi_{1+i} = \xi_1 - \xi_{1+i}$ and $\delta_i^b = b - \xi_{n-i} = \xi_n - \xi_{n-i}$, we have:*

1.

$$\begin{aligned} f(a) &= \theta_1 \ , \\ f(b) &= \theta_m \ . \end{aligned}$$

2.

$$\begin{aligned} f'(a^+) &= d \cdot \frac{\theta_1 - \theta_2}{\delta_1^a} \ , \\ f'(b^-) &= d \cdot \frac{\theta_m - \theta_{m-1}}{\delta_1^b} \ . \end{aligned} \tag{A.15}$$

3.

$$\begin{aligned} f''(a^+) &= \frac{d(d-1)}{\delta_1^a} \left[\frac{\theta_1}{\delta_1^a} - \left(\frac{1}{\delta_1^a} + \frac{1}{\delta_2^a} \right) \theta_2 + \frac{\theta_3}{\delta_2^a} \right] \ , \\ f''(b^-) &= \frac{d(d-1)}{\delta_1^b} \left[\frac{\theta_m}{\delta_1^b} - \left(\frac{1}{\delta_1^b} + \frac{1}{\delta_2^b} \right) \theta_{m-1} + \frac{\theta_{m-2}}{\delta_2^b} \right] \ . \end{aligned} \tag{A.16}$$

Proof. It easily follows by combining Proposition *A*.2.2 and Proposition *A*.2.5 with Proposition *A*.2.3. □

Equations (A.15) and (A.16) are arranged so as to emphasize the symmetry within each pair of expressions: θ_{1+k} and ξ_{1+k} are to a as θ_{m-k} and ξ_{m-k} are to b, respectively, for $k = 0, 1, 2$. Note, however, that $\delta_i^a < 0$ whilst $\delta_i^b > 0$.

A.2.1 A natural spline basis

The combination of (A.15) and (A.16) gives us a parametrization of $\mathbb{S}_{\mathcal{N}}(\Delta)$ in which the first and last coordinates represent the spline slopes at endpoints a and b, respectively.

Proposition A.2.7. *Let Δ be an n-point partition of $[a, b]$, with $n \geq 3$, and let $B_1, B_2, \ldots, B_{n+2}$ be the B-splines for $\mathbb{S}_2^3(\Delta)$. Letting $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n \in \mathbb{R}$ and defining*

$$\begin{aligned} \theta_1 &= \hat{\theta}_2 + \frac{\delta_1^a + \delta_2^a}{3} \hat{\theta}_1 \\ \theta_2 &= \hat{\theta}_2 + \frac{\delta_2^a}{3} \hat{\theta}_1 \\ \theta_i &= \hat{\theta}_{i-1} \ , \ \text{for } i = 3, 4, \ldots, n \ , \\ \theta_{n+1} &= \hat{\theta}_{n-1} + \frac{\delta_2^b}{3} \hat{\theta}_n \\ \theta_{n+2} &= \hat{\theta}_{n-1} + \frac{\delta_1^b + \delta_2^b}{3} \hat{\theta}_n \end{aligned} \tag{A.17}$$

where the δ_i^a, δ_i^b are like in Proposition A.2.6*, it turns out that $f = x \mapsto \sum_{i=1}^{n+2} \theta_i B_i(x) \in \mathbb{S}_{\mathcal{N}}(\Delta)$ with $f'(a^+) = \hat{\theta}_1$ and $f'(b^-) = \hat{\theta}_n$.*

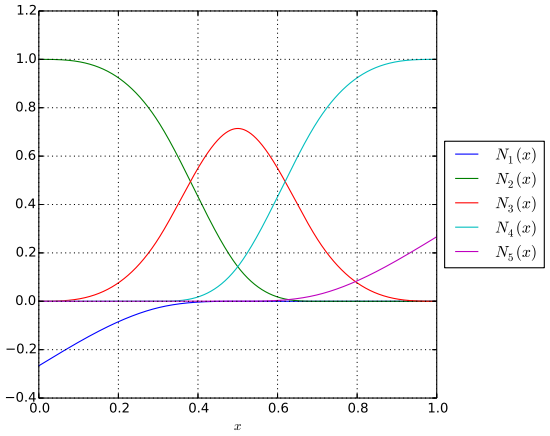


Figure A.2: Natural spline basis for $\Delta = \{0, 0.3, 0.5, 0.7, 1.0\}$ (dimension $n = 5$) induced by parametrization (A.17). Notice that N_3 coincides with B_4 from Figure A.1.

Parametrization (A.17) is derived by enforcing natural boundary conditions (A.3) on the B-spline coordinates. If $\Xi \in \mathcal{M}_{(n+2)\times n}(\mathbb{R})$ is the matrix mapping $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$ to $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{n+2})$, then we can define, for $i = 1, 2, \ldots, n$,

$$N_i(x) = [\Xi^T \mathbf{B}(x)]_i \ , \tag{A.18}$$

where $\mathbf{B}(x) = (B_1(x), B_2(x), \ldots, B_{n+2}(x))$, as a natural spline basis. Figure A.2 depicts basis functions (A.18), while Figure (A.3) shows an instance of a natural spline. As a consequence of $\delta_i^a < 0$ $(i = 1, 2)$, N_1 is the only basis function taking negative values. This somewhat inelegant decision aims at identifying $\hat{\theta}_1$ with $f'(a^+)$, rather than $-f'(a^+)$. Of course, if we wanted all our natural basis functions to be positive, it would suffice to substitute N_1 with $x \mapsto -N_1(x)$.

A.3 Spline smoothness and model regularization

There are a number of spline applications where, in addition to differentiability (local smoothness), global smoothness matters. Differentiability eases optimization to a great extent, but still falls short to overcome model complexity issues. For instance, splines in a regression setting (Eilers and Marx, 1996) may suffer from undesired wiggles, even if they are infinitely differentiable. The process of imposing global smoothness on a non-parametric statistical model (say spline regression) is called *regularization*.

The importance of $\mathbb{S}_2^3(\Delta)$ at this respect roots in physical interpretations. Imagine that an arbitrary $\mathcal{C}^2(a, b)$ function $y(x)$ represents a sufficiently elastic beam. More precisely, we have a function $\sigma : x \in (a, b) \mapsto (x, y(x))$ whose arc

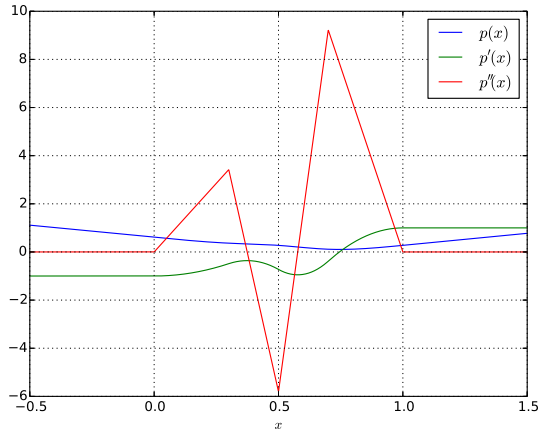


Figure A.3: A randomly generated natural spline p with parametrization (A.17) and knots placed at $\Delta = \{0, 0.3, 0.5, 0.7, 1.0\}$ (dimension $n = 5$). Slope parameters are $\hat{\theta}_1 = p'(0) = -1$ and $\hat{\theta}_5 = p'(1) = 1$, whereas $\hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4 \in [0, 1]$. Linear extrapolation is used to extend p outside $[0, 1]$, according to (A.14).

length between a and x is given by

$$s(x) = \int_a^x \|\sigma'(t)\|_2 \, dt = \int_a^x \sqrt{1 + (y'(t))^2} \, dt \; .$$

If we reparameterize the beam as $\tilde{\sigma} = \sigma \circ s^{-1}$ over $(0, L)$, where $L = s(b)$ is the total length of the beam, we get that $\|\tilde{\sigma}'(s)\|_2 = 1$ for all s, so curvature at arc distance s is simply $\kappa(s) = \|\tilde{\sigma}''(s)\|_2$. The total energy of the beam is proportional to the integral of this quantity over all beam length, that is

$$E \propto \int_0^L \kappa(s)^2 \, ds = \int_a^b \frac{(y''(x))^2}{[1 + (y'(x))^2]^{5/2}} \, dx \; . \tag{A.19}$$

The ideal interpolant function would be the one that, in addition to passing through the specified points, minimized energy (A.19). It turns out that solving for a least energy interpolant, according to this formulation, is no easy task (Horn, 1983). Nonetheless, cubic splines provide a convenient approximation to the solution, according to the following proposition (Shampine et al., 1996, Pollock, 1999).

Proposition A.3.1. *Let $\Delta = \{x_i\}_{i=1}^n$ be a partition of an interval $[a.b]$, with $a < b$, and $\{y_i\}_{i=1}^n$ an arbitrary collection of real values. Let $f \in S_{\mathcal{N}}(\Delta)$ (alternatively, $f \in S_{\alpha,\beta}(\Delta)$, for $\alpha, \beta \in \mathbb{R}$) such that $f(x_i) = y_i$. Spline f minimizes*

$$E(f) = \|f''\|_2^2 = \int_a^b (f''(x))^2 \, dx \; , \tag{A.20}$$

in the sense that, for all $g \in \mathcal{C}^2(a, b)$ passing through points $\{(x_i, y_i)\}_{i=1}^n$ (alternatively, with the additional constraints $g'(a^+) = \alpha$ and $g'(b^-) = \beta$), it verifies $E(f) \leq E(g)$. Moreover, $E(f) = E(g)$ if and only if $f = g$.

Proof. For both natural and clamped boundary conditions, it all comes down to checking $\|(f-g)''\|_2^2 = \|g''\|_2^2 - \|f''\|_2^2$. The uniqueness of the cubic spline interpolant as the minimizer of (A.20) follows from the fact that $\|(f-g)''\|_2 = 0$ yields $f(x) - g(x) = c_1 x + c_2$ for some $c_1, c_2 \in \mathbb{R}$, even if $h \in \mathcal{C}^2(a,b) \mapsto \|h''\|_2$ is not a norm. Finally, since $f(a) = g(a)$ and $f(b) = g(b)$, it follows that $c_1 = c_2 = 0$. □

Energy (A.20) can be easily computed when $f \in \mathbb{S}_2^3(\Delta)$. Let a and b be the left and right endpoints of partition Δ, $N = \dim \mathbb{S}_2^3(\Delta)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)$ be the coordinates of f with respect to the basis splines of $\mathbb{S}_2^3(\Delta)$: $(B_1, B_2, \ldots, B_N)$. We have

$$E(f) = \boldsymbol{\theta}^T \Omega \boldsymbol{\theta} \ , \tag{A.21}$$

where $\Omega \in \mathcal{M}_{N\times N}(\mathbb{R})$ has (i,j)-element

$$\Omega_{ij} = \int_a^b B_i''(x) B_j''(x)\ dx \ . \tag{A.22}$$

Moreover, integral (A.22) can be exactly calculated using Gaussian quadrature formulas like Simpson's rule (Shampine et al., 1996), since the integrand is a piecewise polynomial of degree 2.

Note that (A.20) approximates (A.19) under the assumption that $(y'(x))^2 \ll 1$ for all x in the domain. Also, it is important to remark that the last property holds for both natural and complete boundary conditions; the former, with no additional constraints at all. This is one of the reasons why we usually endow cubic splines with natural boundary conditions.

Going back to regularization, (Pollock, 1999) suggest a regression model $y_i = f(x_i) + \varepsilon_i$, for data points $\{(x_i, y_i)\}_{i=1}^n$ and random variables $\{\varepsilon_i\}_{i=1}^n$, with $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$, minimizing

$$(1-\lambda) \underbrace{\sum_{i=1}^{n} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2}_{\text{data fitting}} + \lambda \underbrace{\int_{x_1}^{x_n} (f''(x))^2\ dx}_{\text{regularization}} \ ,$$

where parameter $\lambda \in [0,1]$ represents the relative importance that is attached to regularization as opposed to data fitting.

Appendix B

Regular variation

This appendix is devoted to summarizing some basic definitions and results of *regular variation*, an analytical property that typically arises in Probability and Statistics (Mikosch, 1999, Soulier, 2009). Probably the most renowned reference on this topic is (Bingham et al., 1987), which includes thorough proofs.

We can compactly state many of the following properties using $\pm$ and $\mp$ notations[1], expressing how different sign choices match one another to produce both 'up' and 'down' alternatives; the former will be considered the first alternative, whereas the latter will take the second place.

In what follows, *measurable* will refer to (real) *Lebesgue*-measurable functions. Also, we shall use the notation $f(x) \sim g(x)$, as $x \to \tau$, whenever $\lim_{x\to\tau} f(x)/g(x) = 1$, and the usual convention $-\int_a^b f(x)\,dx = \int_b^a f(x)\,dx$.

Definition B.0.1 (**Regular variation at ∞**)**.** Let f be a positive, measurable function defined on $[x_0, \infty)$, for some $x_0 \in \mathbb{R}$. We say that f is regularly varying at ∞ with index $\alpha \in \mathbb{R}$, represented by $f \in \mathcal{R}_{\infty,\alpha}$, if

$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} = \lambda^\alpha \ , \tag{B.1}$$

for all $\lambda > 0$.

We now introduce one of the most important properties of regularly varying functions. See (Bingham et al., 1987) for several direct and indirect proofs.

Theorem B.0.1 (**Uniform convergence**)**.** *Let $f \in \mathcal{R}_{\infty,\alpha}$ and $0 < a \leq b < \infty$. The limit (B.1) is uniform[2] in λ on the interval:*

1. *$[a, b]$, if $\alpha = 0$.*
2. *$(0, b]$, if f is bounded on $(0, x]$, for all $x > 0$, and $\alpha > 0$.*

[1] We will assume the notational convention $+\infty = \infty$.

[2] Uniformity in λ on an interval I means that for all $\epsilon > 0$, there exists $x^* > x_0$ (x_0 as in Definition $B.0.1$), such that $x > x^*$ implies

$$\left| \frac{f(\lambda x)}{f(x)} - \lambda^\alpha \right| < \epsilon \ ,$$

regardless of which $\lambda \in I$ is chosen.

3. $[a, \infty)$*, if* $\alpha < 0$*.*

Regular variation can be translated to $-\infty$ in the following manner.

Definition B.0.2 (**Regular variation at** $-\infty$**).** Let f be a positive, measurable function defined on $(-\infty, x_0]$, for some $x_0 \in \mathbb{R}$. We say that f is regularly varying at $-\infty$ with index $\alpha \in \mathbb{R}$, represented by $f \in \mathcal{R}_{-\infty,\alpha}$, if $x \mapsto f(-x) \in \mathcal{R}_{\infty,\alpha}$.

We can also state regular variation at a point z (Mikosch, 1999), both approaching z from left and right.

Definition B.0.3 (**Left regular variation).** Let f be a positive, measurable function defined on $[x_0, z)$, for some $x_0, z \in \mathbb{R}$, $x_0 < z$. We say that f is left regularly varying at z (alternatively, regularly varying at z^-) with index $\alpha \in \mathbb{R}$, represented by $f \in \mathcal{R}_{z^-,\alpha}$, if

$$\lim_{x \to 0^+} \frac{f(z - \lambda x)}{f(z - x)} = \lambda^\alpha \ , \tag{B.2}$$

for all $\lambda > 0$.

Definition B.0.4 (**Right regular variation).** Let f be a positive, measurable function defined on $(z, x_0]$, for some $x_0, z \in \mathbb{R}$, $x_0 > z$. We say that f is right regularly varying at z (alternatively, regularly varying at z^+) with index $\alpha \in \mathbb{R}$, represented by $f \in \mathcal{R}_{z^+,\alpha}$, if

$$\lim_{x \to 0^+} \frac{f(z + \lambda x)}{f(z + x)} = \lim_{x \to 0^-} \frac{f(z - \lambda x)}{f(z - x)} = \lambda^\alpha \ , \tag{B.3}$$

for all $\lambda > 0$.

Pointwise regular variation is connected to regular variation at $\pm\infty$ through the following proposition.

Proposition B.0.2. $f \in \mathcal{R}_{z^\pm,\alpha}$ *if and only if* $x \mapsto f(z - 1/x) \in \mathcal{R}_{\mp\infty,-\alpha}$*.*

Proof. It suffices to see that

$$\lim_{x \to 0^\mp} \frac{f(z - \lambda x)}{f(z - x)} = \lim_{x \to \mp\infty} \frac{f\left(z - 1/(\lambda^{-1} x)\right)}{f(z - 1/x)} \ .$$

□

Proposition B.0.3. *Formally consider* $\tau \in \{\infty, -\infty, z^+, z^-\}$*, that is,* τ *indexes any of the four regular variation scenarios. If* $f \in \mathcal{R}_{\tau,\alpha}$ *and* $f(x) \sim g(x)$ *as* $x \to \tau$*, then* $g \in \mathcal{R}_{\tau,\alpha}$*.*

In all the above regular variation cases, index $\alpha = 0$ plays an important role.

Definition B.0.5 (**Slow variation).** Formally consider $\tau \in \{\infty, -\infty, z^+, z^-\}$, that is, τ indexes any of the four regular variation scenarios. A function $f \in \mathcal{R}_{\tau,0}$ is called slowly varying at τ, represented by $f \in \mathcal{S}_\tau$.

The next two propositions allow to express a regularly varying function as the product of a power function and a slowly varying one, which we shall call its slowly varying part.

Proposition B.0.4. *$f \in \mathcal{R}_{\pm\infty,\alpha}$ if and only if*

$$f(x) = (\pm x)^{\alpha}\ell(x) \ , \tag{B.4}$$

as[3] *$x \to \pm\infty$, for some $\ell \in \mathcal{S}_{\pm\infty}$.*

Proposition B.0.5. *$f \in \mathcal{R}_{z^{\pm},\alpha}$ if and only if*

$$f(x) = [\pm(x-z)]^{\alpha}\ell(x) \ ,$$

as[4] *$x \to z^{\pm}$, for some $\ell \in \mathcal{S}_{z^{\pm}}$.*

By definition of regular variation and using both Proposition $B.0.4$ and Proposition $B.0.5$, we can state the following remark.

Remark B.0.1. For all $\alpha > 0$:

- If $\ell \in \mathcal{S}_{\pm\infty}$, then

$$\lim_{x\to\pm\infty}(\pm x)^{\alpha}\ell(x) = \infty \ , \ \lim_{x\to\pm\infty}(\pm x)^{-\alpha}\ell(x) = 0 \ .$$

- If $\ell \in \mathcal{S}_{z^{\pm}}$, then

$$\lim_{x\to z^{\pm}}[\pm(x-z)]^{-\alpha}\ell(x) = \infty \ , \ \lim_{x\to z^{\pm}}[\pm(x-z)]^{\alpha}\ell(x) = 0 \ .$$

Additionally, slow variation is fully characterized by the next two theorems.

Theorem B.0.2 (Representation theorem at $\pm\infty$). *If $\ell \in \mathcal{S}_{\pm\infty}$, there exists $x_0 \in (0,\infty)$ (alternatively, $x_0 \in (-\infty,0)$) such that, for all $x \geq x_0$ (alternatively, for all $x \leq x_0$),*

$$\ell(x) = c(x)\exp\left(\int_{x_0}^{x}\frac{\eta(t)}{t}\ dt\right) \ , \tag{B.5}$$

for some real functions c and η defined on $[x_0,\infty)$ (alternatively, $(-\infty,x_0]$) and satisfying $\lim_{x\to\pm\infty} c(x) = c^* \in (0,\infty)$ and $\lim_{x\to\pm\infty}\eta(x) = 0$.

Theorem B.0.3 (Representation theorem at $z^{\pm}$). *If $\ell \in \mathcal{S}_{z^{\pm}}$, there exists $x_0 > z$ (alternatively, $x_0 < z$) such that, for all $x \in (z,x_0]$ (alternatively, for all $x \in [x_0,z)$),*

$$\ell(x) = c(x)\exp\left(\int_{x_0}^{x}\frac{\eta(t)}{z-t}\ dt\right) \ , \tag{B.6}$$

for some real functions c and η defined on $x \in (z,x_0]$ (alternatively, $x \in [x_0,z)$) and satisfying $\lim_{x\to z^{\pm}} c(x) = c^* \in (0,\infty)$ and $\lim_{x\to z^{\pm}}\eta(x) = 0$.

Observe that any function ℓ that can be expressed as (B.5) or (B.6) is clearly slowly varying. Also, for every slowly varying function $\ell \in \mathcal{S}_{\tau}$, we can easily construct $\bar{\ell} \in \mathcal{S}_{\tau}$ such that $\ell(x) \sim \bar{\ell}(x)$ as $x \to \tau$ by simply substituting $c(x)$ in (B.5) and (B.6) for its limit as $x \to \tau$, that is c^*.

[3] 'As $x \to +\infty$' actually means 'for all $x \geq x_0$, for some $x_0 \geq 0$'. Similarly, 'as $x \to -\infty$' stands for 'for all $x \leq x_0$, for some $x_0 \leq 0$'.

[4] 'As $x \to z^{-}$' actually means 'for all $x \in [x_0,z)$, for some $x_0 \in \mathbb{R}$'. Similarly, 'as $x \to z^{+}$' stands for 'for all $x \in (z,x_0]$, for some $x_0 \in \mathbb{R}$'.

Definition B.0.6 (Normalized regular variation). If the c function is constant in either (B.5) or (B.6), we shall call the corresponding ℓ normalized and shall write $\ell \in \bar{\mathcal{S}}_\tau$, for $\tau \in \{\infty, -\infty, z^+, z^-\}$. Similarly, a function $f \in \mathcal{R}_{\tau,\alpha}$ shall be called normalized, represented by $f \in \bar{\mathcal{R}}_{\tau,\alpha}$, if its slowly varying part is normalized.

Normalized slowly varying functions have the following interesting property (Soulier, 2009), which is an extension of Remark B.0.1.

Proposition B.0.6. *Let $\alpha > 0$. We have:*

- *$\ell \in \mathcal{S}_{\pm\infty}$ is normalized if and only if both $x \mapsto (\pm x)^\alpha \ell(x)$ and $x \mapsto (\pm x)^{-\alpha}\ell(x)$ are ultimately monotone[5] as $x \to \pm\infty$.*
- *$\ell \in \mathcal{S}_{z^\pm}$ is normalized if and only if both $x \mapsto [\pm(x-z)]^\alpha \ell(x)$ and $x \mapsto [\pm(x-z)]^{-\alpha}\ell(x)$ are ultimately monotone[6] as $x \to z^\pm$.*

Regular variation, in all its different (and equivalent) forms, thus reduces to slow variation. It is time now to present the very first examples of slowly varying functions (at ∞[7]).

Example B.0.1. For any $\beta \in \mathbb{R}$, $x \mapsto (\log x)^\beta$ is slowly varying at ∞.

For continuously differentiable functions, there is an alternative characterization of regular variation (Soulier, 2009).

Proposition B.0.7. *Let ℓ be continuously differentiable. We have that $\ell \in \bar{\mathcal{S}}_{\pm\infty}$ if and only if*

$$\lim_{x\to\pm\infty} \frac{x\ell'(x)}{\ell(x)} = 0 \ . \tag{B.7}$$

Proof. For the 'only if' part, differentiating (B.5), we get

$$\frac{x\ell'(x)}{\ell(x)} = \frac{xc'(x)}{c(x)} + \eta(x) \ .$$

The result follows from c being constant (thus $c'(x) = 0$, for all x) and $\lim_{x\to\pm\infty}\eta(x) = 0$. For the 'if' part, ℓ can be expressed as (B.5), with $\eta(x) = {x\ell'(x)}/{\ell(x)}$ and $c(x) = \ell(x_0)$, for all x some x_0. □

Theorem B.0.4. *Let f be continuously differentiable and let $\alpha \in \mathbb{R}$. We have that $f \in \bar{\mathcal{R}}_{\pm\infty,\alpha}$ if and only if*

$$\lim_{x\to\pm\infty} \frac{xf'(x)}{f(x)} = \alpha \ . \tag{B.8}$$

Proof. The proof comes down to Proposition B.0.7, using Proposition B.0.4 and noting

$$\frac{xf'(x)}{f(x)} = \alpha + \frac{x\ell'(x)}{\ell(x)} \ .$$

□

[5] Meaning monotone in $[x_0, \infty)$ or $(-\infty, x_0]$, respectively.
[6] Meaning monotone in $(z, x_0]$ or $[x_0, z)$, respectively.
[7] Examples at $-\infty$, z^+ and z^- can be derived from this case.

Proposition B.0.8. *A continuously differentiable function f is regularly varying at $z^{\pm}$ with index α if and only if*

$$\lim_{x \to z^{\pm}} \frac{(z-x)f'(x)}{f(x)} = -\alpha \ .$$

Proof. It is a consequence of Theorem $B.0.4$ and Proposition $B.0.2$, letting $g(x) = f(z - 1/x)$:

$$-\alpha = \lim_{x \to \mp\infty} \frac{xg'(x)}{g(x)} = \lim_{x \to \mp\infty} \frac{x^{-1}f'(z-1/x)}{f(z-1/x)} = \lim_{x \to z^{\pm}} \frac{(z-x)f'(x)}{f(x)} \ .$$

□

Integrals of regularly varying functions, under weak assumptions, behave as regularly varying too. Expressing the integrand as a power function times a slowly varying one, the following important result (Mikosch, 1999, McNeil et al., 2005, Soulier, 2009) says we can take the slowly varying factor out of the integral.

Theorem B.0.5 (Karamata's Theorem). *Let $\ell \in S_\infty$ be locally bounded in $[x_0, \infty)$ for some $x_0 \geq 0$.*

1. *If $\alpha > -1$, then*
$$\int_{x_0}^{x} t^{\alpha}\ell(t)\ dt \sim \frac{x^{\alpha+1}}{\alpha+1}\ell(x) \ ,$$
as $x \to \infty$.

2. *If $\alpha < -1$, then*
$$\int_{x}^{\infty} t^{\alpha}\ell(t)\ dt \sim -\frac{x^{\alpha+1}}{\alpha+1}\ell(x) \ ,$$
as $x \to \infty$.

Karamata's theorem is usually stated for slowly varying functions at ∞, but there exist equivalent versions for cases $-\infty$ and $z^{\pm}$ with the same original interpretation.

Corollary B.0.1 (Karamata's Theorem at $-\infty$). *Let $\ell \in S_{-\infty}$ be locally bounded in $(-\infty, x_0]$ for some $x_0 \leq 0$.*

1. *If $\alpha > -1$, then*
$$\int_{x}^{x_0} (-t)^{\alpha}\ell(t)\ dt \sim \frac{(-x)^{\alpha+1}}{\alpha+1}\ell(x) \ ,$$
as $x \to -\infty$.

2. *If $\alpha < -1$, then*
$$\int_{-\infty}^{x} (-t)^{\alpha}\ell(t)\ dt \sim -\frac{(-x)^{\alpha+1}}{\alpha+1}\ell(x) \ ,$$
as $x \to -\infty$.

Proof. It easily follows from a change of variable in Theorem $B.0.5$, noting that, by definition, $\ell \in S_{-\infty}$ if $x \mapsto \ell(-x) \in S_{\infty}$. □

Corollary B.0.2 (Karamata's Theorem at $z^{\pm}$). *Let $\ell \in S_{z^{\pm}}$ be locally bounded in $(z, x_0]$ or in $[x_0, z)$ for some $x_0 > z$ or $x_0 < z$, respectively.*

1. *If $\alpha < -1$, then*

$$\mp \int_{x_0}^{x} [\pm(x-z)]^{\alpha} \ell(t)\ dt \sim -\frac{[\pm(x-z)]^{\alpha+1}}{\alpha+1} \ell(x)\ ,$$

   as $x \to z^{\pm}$.

2. *If $\alpha > -1$, then*

$$\mp \int_{x}^{z} [\pm(x-z)]^{\alpha} \ell(t)\ dt \sim \frac{[\pm(x-z)]^{\alpha+1}}{\alpha+1} \ell(x)\ ,$$

   as $x \to z^{\pm}$.

Proof. This proof is analogous to the one of Corollary $B.0.1$, now using the changes of variable given in Proposition $B.0.2$. □

Next we introduce a concept which is closely related to regular variation.

Definition B.0.7 (Additive regular variation at ∞). A measurable function f defined on $[x_0, \infty)$, for some $x_0 \in \mathbb{R}$, is said to be additively regularly varying at ∞, represented by $f \in \mathcal{R}^{+}_{\infty,\alpha}$, if

$$\lim_{x\to\infty} f(x+\mu) - f(x) = \alpha\mu\ , \tag{B.9}$$

for all $\mu \in \mathbb{R}$.

Expression (B.9) closely resembles that of a '*derivative at infinity*':

$$\lim_{x\to\infty} \frac{f(x+\mu) - f(x)}{\mu} = \alpha\ ,\ \text{if}\ \mu \neq 0\ .$$

Additive regular variation, just like ordinary regular variation, can be studied at $-\infty$.

Definition B.0.8 (Additive regular variation at $-\infty$). A measurable function f defined on $(-\infty, x_0]$, for some $x_0 \in \mathbb{R}$, is said to be additively regularly varying at $-\infty$, represented by $f \in \mathcal{R}^{+}_{-\infty,\alpha}$, if $x \mapsto f(-x) \in \mathcal{R}^{+}_{\infty,-\alpha}$.

Again, additive regular variation is also related to regular variation at $\pm\infty$ by means of the following proposition.

Proposition B.0.9. *$f \in \mathcal{R}^{+}_{\pm\infty,\alpha}$ if and only if $x \mapsto \exp f[\pm\log(\pm x)] \in \mathcal{R}_{\pm\infty,\pm\alpha}$.*

Proof. It suffices to show it is true for $f \in \mathcal{R}^{+}_{\infty,\alpha}$ and then use Definition $B.0.2$ and Definition $B.0.8$. Therefore, letting $g(x) = \exp f(\log x)$, it all comes down to checking

$$\lim_{x\to\infty} f(x+\mu) - f(x) = \log \lim_{x\to\infty} \frac{g(\lambda x)}{g(x)}\ , \tag{B.10}$$

where $\lambda = e^{\mu}$. □

As in Definition $B.0.5$, index $\alpha = 0$ in additive regular variation suggests a slow rate of change.

Definition B.0.9 (Additive slow variation). Formally consider $\tau \in \{\infty, -\infty\}$, that is, τ indexes any of the two additive regular variation scenarios. A function $f \in \mathcal{R}^+_{\tau,0}$ is called additively slowly varying at τ, represented by $f \in \mathcal{S}^+_\tau$.

Just like regularly varying functions could be factorized as a power series times a slowly varying function, additively regularly varying functions can be expressed in terms of their slowly varying counterparts.

Proposition B.0.10. *$f \in \mathcal{R}^+_{\pm\infty,\alpha}$ if and only if*

$$f(x) = \alpha x + \ell(x) \ , \tag{B.11}$$

as $x \to \pm\infty$, for some $\ell \in \mathcal{S}^+_{\pm\infty}$.

Proof. It easily follows from plugging both (B.4) and (B.11) into Proposition $B.0.9$. □

We end this appendix with the additive version of the theorem of uniform convergence.

Proposition B.0.11. *Consider $f \in \mathcal{S}^+_{\pm\infty}$. The limit (B.9) is uniform*[8] *in μ on any interval $[a, b]$.*

Proof. Focusing on the case $+\infty$ ($-\infty$ is analogous) and looking back at the proof of Proposition $B.0.9$, we see that the logarithm function appears both relating constants μ and λ ($\mu = \log \lambda$) and the limits at both sides of equation (B.10). On the one hand, given a closed interval $[a, b]$ for μ, we see that $\lambda \in [\log a, \log b]$, a closed interval. On the other hand, logarithm is continuous at 1. The last two remarks make clear, without any further ϵ-δ reasoning, that the uniformity of f in μ reduces to the uniformity of g in λ (see equation (B.10)), thanks to Theorem $B.0.1$. □

[8] Uniformity in μ on an interval I means that for all $\epsilon > 0$, there exists $x^* > x_0$ (alternatively, $x^* < x_0$), with x_0 as in Definition $B.0.7$ (alternatively, as in Definition $B.0.8$), such that $x > x^*$ (alternatively $x < x^*$) implies $|\ell(x + \mu) - \ell(x)| < \epsilon$, *regardless* of which $\mu \in I$ is chosen.

Appendix C

Quantiles

This appendix aims to remind some well-known properties of quantile functions (McNeil et al., 2005), a concept which, in turn, is one of the cornerstones of copula modelling. As (Embrechts and Hofert, 2013) points out, it is surprisingly unusual to find detailed proofs for these results and, to make matters worse, many of them are wrong or inflated with unnecessary hypotheses. For this reason, we take our time to formalize and tailor them according to our needs.

Through this appendix we will extend, by convention, the usual domain of a univariate CDF, $\mathbb{R}$, to $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ as

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0 \ ,$$

and

$$F(\infty) = \lim_{x \to \infty} F(x) = 1 \ ,$$

assuming, thus, $\{0, 1\} \subset \mathrm{Im}(F)$.

Quantile functions are a special case of *generalized inverses* of CDFs.

Definition C.0.10 (**Quantile function).** Let F be a univariate CDF. We define the quantile function of F, $F^{\leftarrow} : [0, 1] \to \bar{\mathbb{R}}$, as

$$F^{\leftarrow}(\alpha) = \inf\{x \in \bar{\mathbb{R}} : F(x) \geq \alpha\} \ . \tag{C.1}$$

If X is a r.v. with CDF F, we alternatively denote $F^{\leftarrow}$ by $\mathcal{Q}_X$.

From Definition $C.0.10$ we can easily check that

$$F^{\leftarrow}(0) = \inf\{x \in \bar{\mathbb{R}} : F(x) \geq 0\} = \inf \bar{\mathbb{R}} = -\infty \ , \tag{C.2}$$

and $F^{\leftarrow}(\alpha) = \infty$ implies $\alpha = 1$, though the converse is not generally true (it is possible that $F(x) = 1$ for some finite x).

The fact that (C.2) holds for any univariate CDF, including those which have a bounded from below support, may seem a bit odd. Take for example $F(x) = \max\{0, \min\{1, x\}\}$, the CDF of the uniform distribution on $[0, 1]$, for which we would expect $F^{\leftarrow}(0) = 0$. We can justify the use of $\bar{\mathbb{R}}$ in (C.1) just like we allow the domain of a CDF to be $\bar{\mathbb{R}}$. Notwithstanding, quantiles can be sensibly defined to take more informative values at zero (Embrechts and Hofert, 2013), but at the cost of hardening demonstrations. Besides, Embrechts and

Hofert claim that 'from a statistical or practical point of view, the 0-quantile is irrelevant', which is true as regards the results contained in this appendix. Consequently, we follow their advice and adopt Definition $C.0.10$.

Let us introduce the most fundamental properties of quantile functions.

Proposition C.0.12. *Let F be a univariate CDF. The following properties hold:*

1. *$F^{\leftarrow}$ is monotonically increasing.*
2. *For all $\alpha \in [0,1]$, $F(F^{\leftarrow}(\alpha)) \geq \alpha$.*
3. *For all $\alpha \in [0,1]$, if $\alpha \in \mathrm{Im}(F)$, then $F(F^{\leftarrow}(\alpha)) = \alpha$.*
4. *If F is continuous, then $F(F^{\leftarrow}(\alpha)) = \alpha$ for all $\alpha \in [0,1]$.*
5. *For all $x \in \mathbb{R}$ and $\alpha \in [0,1]$, $F(x) \geq \alpha$ if and only if $F^{\leftarrow}(\alpha) \leq x$.*
6. *For all $x \in \mathbb{R}$, $F^{\leftarrow}(F(x)) \leq x$. Moreover, $F^{\leftarrow}(F(x)) < x$ if and only if there exists $x' < x$ such that F is constant on $[x', x]$.*

Proof. Part (1), i.e., $\alpha < \beta$ yielding $F^{\leftarrow}(\alpha) \leq F^{\leftarrow}(\beta)$, directly follows from $\{x \in \bar{\mathbb{R}} : F(x) \geq \beta\} \subset \{x \in \bar{\mathbb{R}} : F(x) \geq \alpha\}$ if $\alpha < \beta$.

To show (2), note that if $F^{\leftarrow}(\alpha) < \infty$, there exists a sequence $\{x_n\}_{n=1}^{\infty}$ such that $x_n \geq F^{\leftarrow}(\alpha)$ for all n and such that $\lim_{n\to\infty} x_n = F^{\leftarrow}(\alpha)$. Since F is right-continuous, $\lim_{n\to\infty} F(x_n) = F(F^{\leftarrow}(\alpha))$ and, as $F(x_n) \geq \alpha$ for all n, we get $F(F^{\leftarrow}(\alpha)) \geq \alpha$. On the other hand, if $F^{\leftarrow}(\alpha) = \infty$, then $\alpha = 1$ and the result follows trivially.

To show (3), suppose that $\alpha \in \mathrm{Im}(F)$. First of all, $F^{-1}(\{\alpha\}) = \{x \in \bar{\mathbb{R}} : F(x) = \alpha\} \subset \{x \in \bar{\mathbb{R}} : F(x) \geq \alpha\}$, so $F^{\leftarrow}(\alpha) \leq \inf F^{-1}(\{\alpha\})$. Secondly, it holds that $F(\inf F^{-1}(\{\alpha\})) = \alpha$, even if $\alpha \in \{0,1\}$ (which are trivial cases), because F is right-continuous. Finally, using (2) and the fact that F is monotonically increasing, we get

$$\alpha \leq F(F^{\leftarrow}(\alpha)) \leq F(\inf F^{-1}(\{\alpha\})) = \alpha \ ,$$

as we wanted to show.

Property (4) is a direct consequence, via (3), of the fact that $(0,1) \subset \mathrm{Im}(F)$ if F is continuous.

As regards (5), the 'only if' part directly follows from Definition $C.0.10$. Conversely, if $F^{\leftarrow}(\alpha) \leq x$, using (2) and being F monotonically increasing, we have $\alpha \leq F(F^{\leftarrow}(\alpha)) \leq F(x)$.

The first part of (6) follows from $x \in \{y \in \bar{\mathbb{R}} : F(y) \geq F(x)\}$. On the other hand, letting $\alpha = F(x)$, if $F^{\leftarrow}(\alpha) < x$, it follows that $F(F^{\leftarrow}(\alpha)) = \alpha = F(x)$, using the fact that $\alpha \in \mathrm{Im}(F)$ and (3). Taking $x' = F^{\leftarrow}(\alpha)$, it holds that $F(x') = F(x)$ and, since F is monotonically increasing, F must be constant on $[x', x]$. Conversely, if $F^{\leftarrow}(F(x)) = x$, by definition of $F^{\leftarrow}$, there cannot exist such an x'. □

Proposition $C.0.12$ leaves a corollary and a very insightful remark.

Corollary C.0.3. *If F is strictly increasing and continuous, $F^{\leftarrow}$ is actually the inverse of F, that is, F^{-1}.*

Proof. Being F strictly increasing guarantees the existence of F^{-1}. Then, using Proposition $C.0.12(4)$, F^{-1} and $F^{\leftarrow}$ must coincide. □

Remark C.0.2. Let $x_0 \in \mathbb{R}$. If $F^{\leftarrow}(F(x_0)) \neq x_0$, then

$$\begin{aligned}&\{x \in \mathbb{R} : F^{\leftarrow}(F(x)) \neq x \text{ and } F(x) = F(x_0)\}\\ &= \begin{cases}(a_0, b_0], & \text{if } b_0 < \infty \text{ and } F \text{ continuous at } b_0\\ (a_0, b_0), & \text{otherwise}\end{cases},\end{aligned}$$

where $a_0 = F^{\leftarrow}(F(x_0)) \in \mathbb{R}$ and $b_0 = \sup\{x \in \bar{\mathbb{R}} : F(x) = F(x_0)\} \in \mathbb{R} \cup \{\infty\}$. This means that $\{x \in \mathbb{R} : F^{\leftarrow}(F(x)) \neq x\}$ can be decomposed as the countable[1] union of pairwise disjoint left-open intervals where F is constant.

We end this appendix introducing the three foremost results on quantile functions.

Proposition C.0.13. *Let $X \sim F$ be a univariate r.v. We have*

$$F^{\leftarrow}(F(X)) = X \text{ a.s. .}$$

Proof. According to Remark $C.0.2$, the set $\{x \in \mathbb{R} : F^{\leftarrow}(F(x)) \neq x\}$ can be decomposed as the countable union of pairwise disjoint left-open intervals $\{I_n\}_{n\in\mathbb{N}}$ where F is constant (say $F(x) = \alpha_n$ for all $x \in I_n$). The probability that X belongs to any of this intervals is zero:

$$\begin{aligned}\mathbb{P}(X \in I_n) &= \begin{cases}F(b_n) - F(a_n), & \text{if } I_n = (a_n, b_n]\\ F(b_n^-) - F(a_n), & \text{if } I_n = (a_n, b_n)\end{cases}\\ &= 0 \ ,\end{aligned}$$

where we have used that F is right-continuous (namely, at a_n) and, thus, $F(a_n) = \alpha_n$. The proof ends by observing

$$\mathbb{P}\left[X \neq F^{\leftarrow}(F(X))\right] = \mathbb{P}\left(X \in \bigcup_{n=1}^{\infty} I_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(X \in I_n\right) = 0 \ .$$

□

Proposition C.0.14. *Let $X \sim F$ be a continuous univariate r.v. Then*

$$F(X) \sim \mathcal{U}[0, 1] \ .$$

Proof. First of all, given $u \in (0, 1)$, since F is monotonically increasing, it necessarily follows that $F^{-1}(\{u\})$ is a closed interval (perhaps containing a single point). Of course, F takes the constant value u on that interval. Besides, being F continuous, the probability of the interval endpoints is zero. Therefore, $\mathbb{P}(X \in F^{-1}(\{u\})) = 0$ and so, given $u \in (0, 1)$,

$$\begin{aligned}\mathbb{P}(F(X) \leq u) &= \mathbb{P}(F(X) < u) + \mathbb{P}(X \in F^{-1}(\{u\}))\\ &= \mathbb{P}(X < F^{\leftarrow}(u)) && \text{(Proposition } C.0.12(5))\\ &= \mathbb{P}(X \leq F^{\leftarrow}(u)) && (X \text{ is a continuous r.v.})\\ &= F(F^{\leftarrow}(u))\\ &= u \ , && \text{(Proposition } C.0.12(4))\end{aligned}$$

as we wanted to show. □

[1] Each interval contains a different rational and $\text{card}(\mathbb{Q}) = \aleph_0$.

Proposition C.0.15. *Let $U \sim \mathcal{U}[0,1]$ and let F be a univariate CDF. We have*

$$F^{\leftarrow}(U) \sim F \ .$$

Proof. It is a direct consequence of Proposition $C.0.12(5)$:

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

□

Appendix D

Implementation details

This annex is devoted to the presentation of several technical details relating to the implementation of the methods described in Chapter 4.

D.1 The bivariate case

In this section we look at the estimation and smoothing procedures described in Section 4.1.3 from an implementation perspective. We will present the expressions for all the gradients involved in the optimization process, the approximations steps to take, numerical tips and computational considerations.

In the following lines we shall suppose that the generators $f_{\boldsymbol{\theta}}$, $F_{\boldsymbol{\theta}}$ and $\phi_{\boldsymbol{\theta}}$ and the copula function $C_{\boldsymbol{\theta}}$ and its density $c_{\boldsymbol{\theta}}$ all relate to the natural spline $g_{\boldsymbol{\theta}}$. Also, we shall denote by $\nabla_{\boldsymbol{\theta}}$ the gradient with respect to the parameters vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{k-1}, \theta_{k+1}, \theta_{k+2}, \ldots, \theta_n)^T \in \mathbb{R}^{n-1}$, when a function, like any of the above mentioned, depends on additional arguments other than $\boldsymbol{\theta}$.

D.1.1 Gradient formulae

The gradient of the natural spline generator $g_{\boldsymbol{\theta}}$ is simply

$$\nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(x) = (N_1(x), N_2(x), \ldots, N_{k-1}(x), N_{k+1}(x), N_{k+2}(x), \ldots, N_n(x))^T \ ,$$

where the natural spline basis functions N_i's are as described in (4.23). The rest of calculations are always the same, regardless of the actual basis or parametrization used.[1] Using the usual construction steps, we get

$$\nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(x) = f_{\boldsymbol{\theta}}(x) \cdot \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}} \left(\sigma^{-1}(x)\right)$$

and differentiating under the integral sign, we have

$$\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(x) = \int_0^x \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(t) \ dt \tag{D.1}$$

and

$$\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}(x) = -\int_x^1 \frac{\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(t)}{(F_{\boldsymbol{\theta}}(t))^2} \ dt \ , \tag{D.2}$$

[1] Remember the missing k-th basis function, which is removed in order to make the model identifiable.

where integration of vector-valued functions must be component-wisely interpreted.

We also need $\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}^{-1}(x)$, which is just a little bit trickier than the previous formulae. By definition,

$$x = \int_{\phi_{\boldsymbol{\theta}}^{-1}(x)}^{1} \frac{1}{F_{\boldsymbol{\theta}}(t)} \, dt \ ,$$

and using implicit differentiation, we have

$$\mathbf{0} = \nabla_{\boldsymbol{\theta}} \left(\int_{\phi_{\boldsymbol{\theta}}^{-1}(x)}^{1} \frac{1}{F_{\boldsymbol{\theta}}(t)} \, dt \right) = -\frac{\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}^{-1}(x)}{F_{\boldsymbol{\theta}} \left(\phi_{\boldsymbol{\theta}}^{-1}(x) \right)} + \nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}} \left(\phi_{\boldsymbol{\theta}}^{-1}(x) \right) \ ,$$

where $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{R}^{n-1}$, to finally obtain

$$\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}^{-1}(x) = F_{\boldsymbol{\theta}} \left(\phi_{\boldsymbol{\theta}}^{-1}(x) \right) \cdot \nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}} \left(\phi_{\boldsymbol{\theta}}^{-1}(x) \right) \ .$$

Special care must be taken when differentiating

$$C_{\boldsymbol{\theta}}(u, v) = \phi_{\boldsymbol{\theta}}^{-1}(\phi_{\boldsymbol{\theta}}(u) + \phi_{\boldsymbol{\theta}}(v)) \ ,$$

with respect to $\boldsymbol{\theta}$.[2] In any case, taking $w = C_{\boldsymbol{\theta}}(u, v)$, the final result is

$$\nabla_{\boldsymbol{\theta}} C_{\boldsymbol{\theta}}(u, v) = F_{\boldsymbol{\theta}}(w) \left[\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}(w) - \nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}(u) - \nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}(v) \right] \ .$$

With all the above ingredients, the gradient of (4.28) is

$$\nabla_{\boldsymbol{\theta}} \log c_{\boldsymbol{\theta}}(u, v) = A + B \ , \tag{D.3}$$

where

$$\begin{aligned} A &= \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}} \left(\sigma^{-1}(w) \right) + w(1 - w) \cdot g'_{\boldsymbol{\theta}} \left(\sigma^{-1}(w) \right) \cdot \nabla_{\boldsymbol{\theta}} C_{\boldsymbol{\theta}}(u, v) \\ B &= \frac{F'_{\boldsymbol{\theta}}(w) \cdot \nabla_{\boldsymbol{\theta}} C_{\boldsymbol{\theta}}(u, v)}{F_{\boldsymbol{\theta}}(w)} + \frac{\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(w)}{F_{\boldsymbol{\theta}}(w)} - \frac{\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(u)}{F_{\boldsymbol{\theta}}(u)} - \frac{\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(v)}{F_{\boldsymbol{\theta}}(v)} \end{aligned}$$

and again $w = C_{\boldsymbol{\theta}}(u, v)$. Note the use of $F'_{\boldsymbol{\theta}}(w)$, instead of $f_{\boldsymbol{\theta}}(w)$, which is meant to emphasize that, whatever we use to approximate $F_{\boldsymbol{\theta}}(w)$, using the derivative of this approximation is numerically more stable than using the original $f_{\boldsymbol{\theta}}(w)$ in the above calculations.

Then, finally, the gradient of the log-likelihood (4.29) is

$$\nabla \log \mathcal{L}(\boldsymbol{\theta} | \mathcal{D}) = \sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} \log c_{\boldsymbol{\theta}}(U_i, V_i) \ .$$

Using the Kendall-based loss function (4.30) instead, the gradient

$$\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(w, z) = \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}} \left(\sigma^{-1}(w) \right) - \frac{\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(z)}{F_{\boldsymbol{\theta}}(z)} \ , \tag{D.4}$$

[2] To avoid confusion, it might be convenient to express it as

$$C(u, v; \boldsymbol{\theta}) = \phi^{-1}(\phi(u; \boldsymbol{\theta}) + \phi(v; \boldsymbol{\theta}); \boldsymbol{\theta}) \ .$$

is much easier to compute than (D.3). In this case, the gradient of the log-likelihood function (4.31) is

$$\nabla \log \mathcal{L}(\boldsymbol{\theta}|\tilde{\mathcal{D}}) = \sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(W_i, Z_i) \ .$$

In either case, as regards the penalty term $E(\boldsymbol{\theta})$, described in (4.34), we simply have

$$\nabla E(\boldsymbol{\theta}) = 2\Omega^* \boldsymbol{\theta} \ .$$

D.1.2 Approximation steps

The functions involved in the optimization problem (4.37) and their derivatives have no closed-form expression. A naive approach to evaluate them results in either numerical errors that will be difficult to overcome or an excessive execution time; measures must be taken to prevent these situations. The techniques employed here might seem a bit tricky, but they are actually commonplace.

The base for all the integral approximations we will make is contained in the following definition.

Definition D.1.1 (Boole's rule). Let $f : \mathbb{R} \to \mathbb{R}$ and let $a, b \in \mathbb{R}$ such that $a < b$. Take $h = (b-a)/4$ and $x_i = a + (i-1)h$, for $i = 1, 2, \ldots, 5$. Define

$$I_a^b(f) = \frac{2h}{45} \left(7f(x_1) + 32f(x_2) + 12f(x_3) + 32f(x_4) + 7f(x_5)\right) \ .$$

The Boole's rule makes the approximation

$$I_a^b(f) \ dx \approx \int_a^b f(x) \ .$$

Should integration be performed at each function evaluation, the optimization process would take too long and become unfeasible. To avoid this, spline interpolation is performed after applying the quadrature in Definition $D.1.1$ on a fine grid. Additionally, when the function is positive, we apply a log-transformation to the quadrature values prior to the interpolation step, increasing its efficacy. The following definitions will make it clear.

Definition D.1.2 (Integral interpolation). Let $f : \mathbb{R} \to \mathbb{R}$. Let $a, b \in \mathbb{R}$ such that $a < b$ and $m \in \mathbb{N}$, with $m > 2$. Take $h = (b-a)/m$ and $x_k = a + kh$, for $k = 0, 1, \ldots, m$. Define the sequence of partial sums $\{q_k\}_{k=0}^m$ as

$$q_{k+1} = q_k + I_{x_k}^{x_{k+1}}(f) \ ,$$

beginning with $q_0 = 0$. We define $\hat{Q}_{[a,b]}^m\{f\} : [a, b] \to \mathbb{R}$ as the cubic spline interpolating the set of points $\{(x_k, q_k)\}_{k=1}^m$ and having flat continuations outside $[a + h, b]$. We can make the approximation

$$\hat{Q}_{[a,b]}^m\{f\}(x) \approx \int_a^x f(x) \ dx \ ,$$

for $x \in [a + h, b]$.

Definition D.1.3 (**Backwards integral interpolation).** Let $f : \mathbb{R} \to \mathbb{R}$. Let $a, b \in \mathbb{R}$ such that $a < b$ and $m \in \mathbb{N}$, with $m > 2$. Take $h = (b-a)/m$ and $x_k = b - (m-k+1)h$, for $k = 1, 2, \ldots, m+1$. Define the sequence of partial sums $\{q_k\}_{k=1}^{m+1}$ as

$$q_k = q_{k+1} + I_{x_k}^{x_{k+1}}(f) \ ,$$

beginning with $q_{m+1} = 0$. We define $\check{Q}_{[a,b]}^m\{f\} : [a, b] \to \mathbb{R}$ as the cubic spline interpolating the set of points $\{(x_k, q_k)\}_{k=1}^m$ and having flat continuations outside $[a, b-h]$. We can make the approximation

$$\check{Q}_{[a,b]}^m\{f\}(x) \approx \int_x^b f(x)\ dx \ ,$$

for $x \in [a, b-h]$.

When the function to be integrated is positive and is known to range over several orders of magnitude, it is advisable to interpolate log-scale values.

Definition D.1.4 (**Log-scale integral interpolation).** Let $f : \mathbb{R} \to (0, \infty)$. Let $a, b \in \mathbb{R}$ such that $a < b$ and $m \in \mathbb{N}$, with $m > 2$. Take $h = (b-a)/m$ and $x_k = a + kh$, for $k = 0, 1, \ldots, m$. Define the sequence of partial sums $\{q_k\}_{k=0}^m$ as

$$q_{k+1} = q_k + I_{x_k}^{x_{k+1}}(f) \ ,$$

beginning with $q_0 = 0$. We define $\hat{L}_{[a,b]}^m\{f\} : [a, b] \to (0, \infty)$ as the cubic spline interpolating the set of points $\{(x_k, \log q_k)\}_{k=1}^m$ and having flat continuations outside $[a+h, b]$. We can make the approximation

$$\hat{L}_{[a,b]}^m\{f\}(x) \approx \log \int_a^x f(x)\ dx \ ,$$

for $x \in [a+h, b]$.

Definition D.1.5 (**Backwards log-scale integral interpolation).** Let $f : \mathbb{R} \to (0, \infty)$. Let $a, b \in \mathbb{R}$ such that $a < b$ and $m \in \mathbb{N}$, with $m > 2$. Take $h = (b-a)/m$ and $x_k = b - (m-k+1)h$, for $k = 1, 2, \ldots, m+1$. Define the sequence of partial sums $\{q_k\}_{k=1}^{m+1}$ as

$$q_k = q_{k+1} + I_{x_k}^{x_{k+1}}(f) \ ,$$

beginning with $q_{m+1} = 0$. We define $\check{L}_{[a,b]}^m\{f\} : [a, b] \to (0, \infty)$ as the cubic spline interpolating the set of points $\{(x_k, \log q_k)\}_{k=1}^m$ and having flat continuations outside $[a, b-h]$. We can make the approximation

$$\check{L}_{[a,b]}^m\{f\}(x) \approx \log \int_x^b f(x)\ dx \ ,$$

for $x \in [a, b-h]$.

We shall extend component-wisely the definition of the all the above operators to vector-valued functions. Letting T represent any integral operator and $\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_d(x))$, we define

$$T\{\mathbf{f}\}(x) = (T\{f_1\}(x), T\{f_2\}(x), \ldots, T\{f_d\}(x)) \ .$$

Next, we present explicit formulae for the approximation of $F_{\boldsymbol{\theta}}(x)$, $\phi_{\boldsymbol{\theta}}(x)$ and their gradients and derivatives over a sufficiently large interval $I = [\epsilon, 1-\epsilon] \subset [0,1]$, where ϵ represents a really small positive quantity like $\epsilon = 10^{-10}$. Following the recommendation in (Lambert, 2007), we stabilize all integrands by means of a change of variable, namely $t = \sigma(s)$, and eventually integrate over $\sigma^{-1}(I) = [\sigma^{-1}(\epsilon), \sigma^{-1}(1-\epsilon)] \subset \mathbb{R}$.

- To approximate $F_{\boldsymbol{\theta}}$ in I:

$$\begin{aligned} F_{\boldsymbol{\theta}}(x) &= \int_{-\infty}^{\sigma^{-1}(x)} \exp g_{\boldsymbol{\theta}}(s)\ d\sigma(s) \\ &= F_{\boldsymbol{\theta}}(\epsilon) + \int_{\sigma^{-1}(\epsilon)}^{\sigma^{-1}(x)} \exp g_{\boldsymbol{\theta}}(s)\ d\sigma(s) \\ &\approx \int_{\sigma^{-1}(\epsilon)}^{\sigma^{-1}(x)} \exp g_{\boldsymbol{\theta}}(s)\ d\sigma(s) \\ &\approx \exp\left[\hat{L}^m_{\sigma^{-1}(I)}\{\rho\}\left(\sigma^{-1}(x)\right)\right] = \tilde{F}_{\boldsymbol{\theta}}(x) \end{aligned} , \tag{D.5}$$

  where

$$\rho(s) = \sigma'(s) \cdot \exp g_{\boldsymbol{\theta}}(s)\ . \tag{D.6}$$

- To approximate $F'_{\boldsymbol{\theta}}$ in I, do not use $f_{\boldsymbol{\theta}}$, but the derivative of the estimate $\tilde{F}_{\boldsymbol{\theta}}$. Let $p(x) = \hat{L}^m_{\sigma^{-1}(I)}\{\rho\}(x) = \log \tilde{F}_{\boldsymbol{\theta}}(\sigma(x))$, where ρ is as defined in (D.6). Then, we can use

$$F'_{\boldsymbol{\theta}}(x) \approx \frac{\tilde{F}_{\boldsymbol{\theta}}(x) \cdot p'\left(\sigma^{-1}(x)\right)}{x(1-x)}\ .$$

- To approximate $\phi_{\boldsymbol{\theta}}$ in I, we use the approximation for $\tilde{F}_{\boldsymbol{\theta}}$ in (D.5):

$$\begin{aligned} \phi_{\boldsymbol{\theta}}(x) &= \int_{\sigma^{-1}(x)}^{\infty} \frac{1}{F_{\boldsymbol{\theta}}(\sigma(s))}\ d\sigma(s) \\ &= \phi_{\boldsymbol{\theta}}(1-\epsilon) + \int_{\sigma^{-1}(x)}^{\sigma^{-1}(1-\epsilon)} \frac{1}{F_{\boldsymbol{\theta}}(\sigma(s))}\ d\sigma(s) \\ &\approx \int_{\sigma^{-1}(x)}^{\sigma^{-1}(1-\epsilon)} \frac{1}{F_{\boldsymbol{\theta}}(\sigma(s))}\ d\sigma(s) \\ &\approx \int_{\sigma^{-1}(x)}^{\sigma^{-1}(1-\epsilon)} \frac{1}{\tilde{F}_{\boldsymbol{\theta}}(\sigma(s))}\ d\sigma(s) \\ &\approx \exp\left[\check{L}^m_{\sigma^{-1}(I)}\{\rho\}\left(\sigma^{-1}(x)\right)\right] = \tilde{\phi}_{\boldsymbol{\theta}}(x) \end{aligned} , $$

  where

$$\rho(s) = \frac{\sigma'(s)}{\tilde{F}_{\boldsymbol{\theta}}(\sigma(s))}\ . \tag{D.7}$$

- To approximate $\phi^{-1}_{\boldsymbol{\theta}}$, let $\{(x_k, y_k)\}_{k=1}^m$ be the points that give rise to the interpolating spline $p(x) = \check{L}^m_{\sigma^{-1}(I)}\{\rho\}(x) = \log \tilde{\phi}_{\boldsymbol{\theta}}(\sigma(x))$, where ρ is as

defined in (D.7). Let q be the cubic spline interpolating $\{(\bar{y}_k, \bar{x}_k)\}_{k=1}^m$, where $\bar{x}_k = x_{m-k+1}$ and $\bar{y}_k = y_{m-k+1}$, and thus satisfying $q(x) \approx p^{-1}(x)$.[3] Then,

$$\phi_{\boldsymbol{\theta}}^{-1}(x) \approx \sigma\left(q(\log x)\right) \ .$$

- To approximate $\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}$ in I:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(x) &= \int_{-\infty}^{\sigma^{-1}(x)} \exp g_{\boldsymbol{\theta}}(s) \cdot \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(s) \ d\sigma(s) \\
&= \nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(1-\epsilon) + \int_{\sigma^{-1}(\epsilon)}^{\sigma^{-1}(x)} \exp g_{\boldsymbol{\theta}}(s) \cdot \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(s) \ d\sigma(s) \\
&\approx \int_{\sigma^{-1}(\epsilon)}^{\sigma^{-1}(x)} \exp g_{\boldsymbol{\theta}}(s) \cdot \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(s) \ d\sigma(s) \\
&\approx \hat{Q}^m_{\sigma^{-1}(I)}\{\boldsymbol{\rho}\}\left(\sigma^{-1}(x)\right) = \tilde{\nabla}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(x)
\end{aligned} \quad , \tag{D.8}$$

  where

$$\boldsymbol{\rho}(s) = \sigma'(s) \cdot \exp g_{\boldsymbol{\theta}}(s) \cdot \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(s) \ .$$

- To approximate $\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}$ in I, we use the approximation $\tilde{\nabla}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}$ in (D.8).

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}(x) &= \int_{\sigma^{-1}(x)}^{\infty} \frac{1}{[F_{\boldsymbol{\theta}}(\sigma(s))]^2} \cdot \nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(\sigma(s)) \ d\sigma(s) \\
&= \nabla_{\boldsymbol{\theta}} \phi_{\boldsymbol{\theta}}(1-\epsilon) + \int_{\sigma^{-1}(x)}^{\sigma^{-1}(1-\epsilon)} \frac{1}{[F_{\boldsymbol{\theta}}(\sigma(s))]^2} \cdot \nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(\sigma(s)) \ d\sigma(s) \\
&\approx \int_{\sigma^{-1}(x)}^{\sigma^{-1}(1-\epsilon)} \frac{1}{[F_{\boldsymbol{\theta}}(\sigma(s))]^2} \cdot \nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(\sigma(s)) \ d\sigma(s) \\
&\approx \int_{\sigma^{-1}(x)}^{\sigma^{-1}(1-\epsilon)} \frac{1}{[\tilde{F}_{\boldsymbol{\theta}}(\sigma(s))]^2} \cdot \tilde{\nabla}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(\sigma(s)) \ d\sigma(s) \\
&\approx \check{Q}^m_{\sigma^{-1}(I)}\{\boldsymbol{\rho}\}\left(\sigma^{-1}(x)\right)
\end{aligned} \quad ,$$

  where

$$\boldsymbol{\rho}(s) = \frac{\sigma'(s)}{[\tilde{F}_{\boldsymbol{\theta}}(\sigma(s))]^2} \cdot \tilde{\nabla}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(\sigma(s)) \ .$$

- If we denote the original estimate of the copula CDF by $\tilde{C}_{\boldsymbol{\theta}}$, arising from all the previous approximations, we can improve it with

$$C_{\boldsymbol{\theta}}(u,v) \approx \min\left\{1-\epsilon, \max\left\{\tilde{C}_{\boldsymbol{\theta}}(u,v), \epsilon\right\}\right\} \ .$$

With all the above approximations, we are ready to run a standard optimization algorithm relying on function and gradient evaluations to solve (4.37), regardless of the chosen loss-function. Despite the apparent complexity and instability-proneness of the previous approximation steps, high accuracy is achieved, as we were able to check by comparing with finite-difference estimates of the gradients involved.

[3] Since $\phi_{\boldsymbol{\theta}}$ is strictly decreasing, we have $y_1 > y_2 > \ldots > y_m$.

The number of points m used for integrals interpolation must be sufficiently large to ensure the accuracy of calculations over the whole $\sigma^{-1}(I)$. However, a very large m will slow down computations. In our experience, $m = 500$ provides good approximations if combined with $\epsilon = 10^{-10}$, while keeping execution times under control.

D.1.3 Practical tips

In the above calculations, the finite precision arithmetic of computers causes some minor numerical errors that, if not properly addressed, may harm the optimization process. Namely, during the integration steps, some unexpected 0's may arise, despite $F_{\boldsymbol{\theta}}(x)$ and $\phi_{\boldsymbol{\theta}}(x)$ being always positive for $x \in (0, 1)$. For instance, it is not unusual to have $\exp g_{\boldsymbol{\theta}}(s) \approx 0$ or $\sigma'(s) \approx 0$ for sufficiently negative s's, in which case the computer might erroneously return 0. Consequently, the logarithmic transformations we use before interpolation in (D.1.4) and (D.1.5) return $-\infty$'s, which obviously cannot be interpolated. This values must be removed from the interpolation grid, along with their corresponding abscissae. Fortunately, this step does not affect the overall accuracy of the estimates.

D.1.4 Computational considerations

Fixed the number of knots n, the complexity of the estimation process is $\mathcal{O}(N)$, where N is the sample size, due to the log-likelihood evaluations. On the other hand, if N remains constant, the complexity is $\mathcal{O}(n)$, where n is the number of spline knots used, which equals the number of basis functions plus one.

Specially for small sample sizes, the bottleneck of any gradient-based optimization algorithm applied to our problem is the construction of the estimate functions. Apart from estimating the latent functions, each gradient comprises $n - 1$ partial derivatives. Each one involves, besides an interpolation step, m runs of the 5-point quadrature rule (D.1.1). Notwithstanding, note that there are *far* fewer function evaluations when using the lighter gradient (D.4).

We have successfully tried both gradient-based and gradient-free optimization algorithms: L-BFGS (a light version of the BFGS quasi-Newton algorithm used in (Hernández-Lobato and Suárez, 2011)) and Nelder-Mead, respectively. In addition to easily handling the non-negative slope constraint $\theta_1 \geq 0$ (4.36), both converge without needing a decent initial guess at the solution. This contrasts with the recommendation in (Hernández-Lobato and Suárez, 2011) of using some parts of (Lambert, 2007) to set the initial guess. As a matter of fact, our implementation using the NLopt library[4] in Julia (Bezanson et al., 2012) demonstrates Hernández-Lobato and Suárez's method converges by itself alone, taking as initial estimate $\boldsymbol{\theta}_0 = \mathbf{0}_{n-1}$, the product/independence copula.

The gradient-based algorithm rapidly converges. Most times, it barely requires from 10 to 20 *heavy* iterations for the log-likelihood target value to stabilize up to a reasonable tolerance, taking only a few seconds. By contrast, the gradient-free algorithm struggles to converge and usually takes far longer, needing a huge number of light iterations.

[4] Steven G. Johnson, *The NLopt nonlinear-optimization package*, available at `http://ab-initio.mit.edu/nlopt`

D.2 The conditional case

We now present further implementation details on the proposal we made in Section 4.2. Much of the new material actually builds upon the previous section. To begin with, all gradient formulae from Section D.1.1 also apply in the conditional case, just replacing the parameter vector $\boldsymbol{\theta}$ with the parameter matrix Θ and adding an extra argument for the covariate w. For instance, (D.1) becomes

$$\nabla_\Theta F_\Theta(x, w) = \int_0^x \nabla_\Theta f_\Theta(t, w)\ dt\ , \tag{D.9}$$

where $F_\Theta(*, w)$ and $f_\Theta(*, w)$ play the role of F_w and f_w in (4.40), respectively.

D.2.1 Approximation steps

The most relevant implementation detail concerns the approximation of the bivariate functions involved in the estimation of the model, namely those in equation (4.45). In the non-conditional case, we used cubic spline interpolation to approximate those functions; now we shall use bi-quadratic splines.

Unfortunately, we cannot directly apply bivariate versions of the interpolation operators described in the previous section. As mentioned in Section D.1.3, due to the finite precision arithmetic of computers, some intermediate calculations that happen to be inaccurate have to be removed for the sake of the entire estimation process. Since the x-components to be removed vary from one covariate w to another, the resulting grid is not rectangular, which makes interpolation just not feasible.

To overcome this problem, we propose to perform an intermediate unidimensional interpolation step. Let $0 = w_1 < w_2 < \ldots < w_k = 1$ be k equidistant knots in the covariate w-space and let the interpolation grid $a = x_1 < x_2 < \ldots < x_m = b$ in $\mathbb{R}$. Consider, for instance, the approximation of $x \mapsto F_\Theta(x, w_j)$ according to (D.5), i.e., $x \mapsto \tilde{F}_\Theta(x, w_j)$. We can approximate the function F_Θ in both variables, x and w, by means of a bi-quadratic spline $p(x, w)$ interpolating the values $y_{ij} = \log \tilde{F}_\Theta(\sigma(x_i), w_j)$ over the grid $\{(x_i, w_j) : i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, k\}$ and finally defining $\bar{F}_\Theta(x, w) = \exp p(\sigma^{-1}(x), w)$. The same process works for the rest of relevant functions and derivatives.

A different, but related, problem concerns the inversion of the Archimedean generator $x \mapsto \phi_\Theta(x, w)$. Because the range of each generator varies from one covariate w to another, we need to find a common domain for the inverse generators $y \mapsto \phi_\Theta^{-1}(y, w)$. We propose the following steps:

1. Consider the values $y_{ij} = \log \tilde{\phi}_\Theta(\sigma(x_i), w_j)$ used to obtain the final estimate $\bar{\phi}_\Theta(x, w)$.

2. Fixed w_j, build a cubic spline $p_j(y)$ interpolating the values x_i over the grid y_{ij}. This way, $p_j(y) \approx \sigma^{-1}(\phi_\Theta^{-1}(\exp y, w_j))$.

3. Take m equidistant knots $\{y_k\}_{k=1}^m$ between $y_1 = \min\{y_{ij}\}$ and $y_m = \max\{y_{ij}\}$.

4. Define $x_{ij} = p_j(y_i)$, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, k$.

5. Build a bi-quadratic spline $p(y, w)$ interpolating the values x_{ij} over the grid $\{(y_i, w_j) : i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, k\}$.

6. Finally define $\bar{\phi}_{\Theta}^{-1}(y, w) = \sigma\left[p(\log y, w)\right]$.

In order to get accurate approximations, our experience shows we need about $k = 100$ covariate values. For the rest of model settings, we can use $m = 500$ and $\epsilon = 10^{-10}$, as in the bivariate case. We successfully tried to alleviate the high computational costs and make up for the large number of function evaluations by employing bi-quadratic splines instead of bi-cubic splines.

D.2.2 Computational considerations

The estimation and smoothing process is computationally expensive, largely due to the intermediate approximations we have just described. Given a parameter configuration, the complexity of the algorithm is again $\mathcal{O}(N)$, where N is the sample size. However, this time the number of parameters grows really fast with the number of basis functions n and m in the x and w axis, respectively. This fact has an impact on both the estimation stage itself and the preceding approximation stage. Namely, fixing n, the complexity is $\mathcal{O}(m)$; fixing m, we have $\mathcal{O}(n)$.

Contrary to the bivariate case, a gradient-free optimization algorithm like Nelder-Mead did not work for us. The gradient-based L-BFGS algorithm did work, also without a good initial guess at the solution. Specifically, convergence is accomplished if we set all the entries of parameter matrix Θ to 0, which is equivalent to considering $\phi_w(x) = -\log x$ (independence copula) for all $w \in [0, 1]$.

Bibliography

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67(2):445–53.

Acar, E. F., Genest, C., and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90.

Bedford, T. and Cooke, R. M. (2002). Vines - A new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068.

Bezanson, J., Karpinski, S., Shah, V. B., and Edelman, A. (2012). Julia: A Fast Dynamic Language for Technical Computing. *arXiv:1209.5145*, pages 1–27.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*, volume 27. Cambridge University Press.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). Copulas for Finance - A Reading Guide and Some Applications. *SSRN Electronic Journal*.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11:365–385.

Charpentier, A., Fermanian, J.-d., and Scaillet, O. (2006). The Estimation of Copulas: Theory and Practice.

Charpentier, A., Fougères, A. L., Genest, C., and Nešlehová, J. G. (2014). Multivariate Archimax copulas. *Journal of Multivariate Analysis*, 126:118–136.

Charpentier, A. and Segers, J. (2009). Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis*, 100(7):1521–1537.

de Boor, C. (1972). On calculating with B-splines. *Journal of Approximation Theory*, 6(1):50–62.

de Boor, C. and Daniel, J. W. (1974). Splines with Nonnegative B-Spline Coefficients. *Mathematics of Computation*, 28(126):565.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78.

Dovgoshey, O., Martio, O., Ryazanov, V., and Vuorinen, M. (2006). The cantor function. *Expositiones Mathematicae*, 24(1):1–37.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B -splines and penalties. *Statistical Science*, 11(2):89–121.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432.

Embrechts, P., Lindskog, F., and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In Rachev, S., editor, *Handbook of Heavy Tailed Distributions in Finance*, chapter 8, pages 329–384. Elsevier.

Gagliardini, P. and Gouriéroux, C. (2007). An efficient nonparametric estimator for models with nonlinear dependence. *Journal of Econometrics*, 137(1):189–229.

Genest, C. and Rémillard, B. (2006). Discussion of "Copulas: Tales and facts", by Thomas Mikosch. *Extremes*, 9(1):27–36.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043.

Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their applications. *Computational Statistics & Data Analysis*, 55(5):1919–1932.

Gudendorf, G. and Segers, J. (2010). Extreme-Value Copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, volume 198 of *Lecture Notes in Statistics*, chapter 6, pages 127–145. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hernández-Lobato, J. M. and Suárez, A. (2011). Semiparametric bivariate Archimedean copulas. *Computational Statistics and Data Analysis*, 55:2038–2058.

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

Horn, B. K. P. (1983). The Curve of Least Energy. *ACM Transactions on Mathematical Software*, 9(4):441–460.

Hua, L. and Joe, H. (2013). Intermediate Tail Dependence: A Review and Some New Results. In *Stochastic Orders in Reliability and Risk: In Honor of Professor Moshe Shaked*, pages 291–311.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag, New York.

Lambert, P. (2007). Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics and Data Analysis*, 51:6307–6320.

Lambert, P. (2014). Spline approximations to conditional Archimedean copula. *Stat*, 3(1):200–217.

Lopez-Paz, D., Hernández-Lobato, J. M., and Ghahramani, Z. (2013). Gaussian Process Vine Copulas for Multivariate Dependence. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA.

Lyche, T. and Morken, K. (2002). *Spline Methods*.

Lyche, T. and Schumaker, L. L. (1975). Local spline approximation methods. *Journal of Approximation Theory*, 15(4):294–325.

Mao, W. and Zhao, L. H. (2003). Free-knot polynomial splines with confidence intervals. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65(4):901–919.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Mikosch, T. (1999). Regular variation, subexponentiality and their applications in probability theory. Technical report, Eurandom Institute, Eindhoven.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9(1):3–20.

Nelsen, R. B. (2003). Properties and applications of copulas: A brief survey. In Dhaene, J., Kolev, N., and Morettin, P., editors, *Proceedings of the First Brazilian Conference on Statistical Modelling in Insurance and Finance*, volume 3, pages 1–18, São Paulo.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag, New York.

Patton, A. J. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, 47(2):527–556.

Pollock, D. S. G. (1999). Smoothing with Cubic Splines. In *Handbook of Time Series Analysis, Signal Processing and Dynamics*, chapter 11, pages 293–322. Elsevier.

Schmitz, V. (2003). *Copulas and Stochastic Processes*. Phd, Aachen University.

Schumaker, L. L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press, New York, 3rd edition.

Shampine, L. F., Allen, R. C., and Pruess, S. (1996). *Fundamentals of numerical computing*, volume 32. John Wiley & Sons, New York.

Sklar, A. (1973). Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika*, 9(6):449–460.

Soulier, P. (2009). *Some applications of regular variation in probability and statistics*. Instituto Venezolano de Investigaciones Científicas, Caracas.

Vandenhende, F. and Lambert, P. (2005). Local dependence estimation using semiparametric archimedean copulas. *Canadian Journal of Statistics*, 33(3):377–388.

Wood, S. N. and Augustin, N. H. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, 157(2-3):157–177.

Zhang, H. and Ding, F. (2013). On the Kronecker products and their applications. *Journal of Applied Mathematics*, 2013.

QVID VLTRA FACIAM?