# Aggregation-Tree Copula Models in Extreme-Value Theory

Sebastian Pacheco

Department of Mathematics and Statistics

McGill University, Montréal

June, 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Arts

# Abstract

This thesis gives an overview of an approach originally described by Arbenz et al. (2012) and expanded upon in Côté & Genest (2015) to model high-dimensional data using low-dimensional copulas. This modeling approach, which is based on aggregation trees, relies on a conditional independence assumption. An alternative assumption of similar nature is suggested, which is equivalent but simpler to verify than those used by Arbenz et al. (2012) and Côté & Genest (2015). A framework is also suggested to use these models for extreme-value analysis. The framework is illustrated using a hydrometric dataset collected from various monitoring stations scattered in the province of Québec. The necessary conditions and assumptions for an aggregation-tree model to generate multivariate extreme-value distributions are also explored.

# Résumé

Ce mémoire donne une vue d'ensemble d'une approche préconisée par Arbenz, Hummel & Mainik (2012) et développée par Côté & Genest (2015) aux fins de modélisation de données de grande dimension au moyen de copules de basse dimension. Cette approche, qui s'appuie sur des structures d'agrégation arborescentes, repose sur un postulat d'indépendance conditionnelle. Un nouveau postulat de nature semblable est proposé qui s'avère équivalent mais plus simple à vérifier que ceux employés par Arbenz et al. (2012) et Côté & Genest (2015). Un cadre est en outre suggéré pour l'utilisation de tels modèles pour l'analyse des valeurs extrêmes. Son emploi est illustré à l'aide d'un jeu de données hydrométriques recueillies par diverses stations de surveillance réparties sur le territoire du Québec. Les conditions et postulats requis pour qu'un modèle d'agrégation arborescent génère des lois multivariées de valeurs extrêmes sont aussi explorés.

# Acknowledgements

I am grateful to my supervisors, Professors Christian Genest and Johanna Nešlehová, for their guidance and support in relation with this research project. Their immense knowledge of copula models and technical writing, which they kindly shared with me, was very valuable throughout this project. I thank them for this incredible learning experience.

Additionally, I would like to thank my family for supporting me throughout my studies. I am grateful to them for encouraging me to pursue my passion for mathematics in the Master's program.

Funding in support of this work was provided through grants to Professors Genest and Nešlehová by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chairs Program. This funding is gratefully acknowledged.

# Preface and Author Contributions

The goal of this project was to explore the applications of aggregation-tree models as seen in the work of Arbenz et al. (2012) and Côté & Genest (2015) in extreme-value theory.

Arbenz et al. (2012) presented a proof to show that the models generate unique and well-defined multivariate distributions under certain conditional independence assumptions. However, this proof was somewhat incomplete. In Chapter 2, with the help of my supervisors, I produced a more complete proof than the one from Arbenz et al. (2012). Through this proof, I found another formulation of the conditional independence assumption used for these models and proved its equivalence to the formulations used in both of the previously mentioned articles. Additionally, my supervisors suggested a notation framework for aggregation-tree models which was used in various proofs in this chapter and throughout the thesis.

In Chapter 5, we present a framework to fit aggregation-tree models for use in the analysis of extreme-value data. We proposed the use $F$-madograms to select a tree structure for the model, as was recommended by Prof. Nešlehová. In Chapter 6, an aggregation-tree model was fitted with the framework presented in Chapter 5 to a dataset consisting of mean monthly discharge data from hydrometric stations in the province of Québec to illustrate the framework's potential use in hydrometry. I received guidance from my supervisors, as they recommended various tests to use for validation and the use of certain copulas for the model.

In Chapter 7, I present partial results relating to the conditions under which a multivariate extreme-value distribution can be generated using aggregation-tree models. I

propose an additional assumption for these models that would ensure that a multivariate extreme-value distribution is obtained. My supervisors helped greatly to find and correct various mistakes in my proofs in this chapter and pointed out limitations for my suggested assumption. My supervisors also edited most of the chapters of this thesis to help improve the formatting.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This chapter is meant to present the prerequisite theory to understand the concept of copula as a tool for modeling dependence among continuous random variables. A seminal result by Sklar (1959) is the theoretical basis for this approach. Indeed, Abe Sklar (1925–2020) showed how copulas can be used to uniquely characterize the dependence structure between sets of continuous random variables, regardless of their marginal distributions. For simplicity, we shall focus throughout this thesis on continuous random variables, though similar results exist for collections of discrete random variables or combinations of continuous and discrete random variables; see, e.g., Genest & Nešlehová (2007).

We will also introduce a class of models due to Arbenz et al. (2012) for the purpose of constructing copula models for large numbers of variables. We intend to examine the necessary conditional independence assumptions that these models rely on to obtain a unique and well-defined joint distribution. We will also provide a simpler formulation of this assumption which is equivalent to, but simpler to check than, those used by Arbenz et al. (2012) and by Côté & Genest (2015). Moreover, we will examine some of the properties of these models in the context of extreme-value theory.

## 1.1 Copulas and copula models

Many of the results and definitions presented here are outlined, e.g., in the books by Nelsen (2006) and McNeil et al. (2015). We begin with the standard definition of a copula and an overview of the properties of these functions.

**Definition** (copula)**.** Given a $d$-dimensional random vector $U = (U_1, \ldots, U_d)$ with uniform marginal distributions on $[0, 1]$, we call the distribution function $C$ of $U$ a copula and, for convenience, we restrict its domain to the set $[0, 1]^d$.

By definition, if $C$ is a copula, then it satisfies the following conditions:

(i) For arbitrary $i \in \{1, \ldots, d\}$ and $u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_d \in (0, 1)$, the map $u_i \mapsto C(u_1, \ldots, u_i, \ldots, u_d)$ is non-decreasing.

(ii) The margins of $C$ are all uniform on $[0, 1]$, i.e., for any $i \in \{1, \ldots, d\}$, $C(u_1, \ldots, u_d) = u_i$ if $u_j = 1$ for all $j \neq i$ and $u_i \in [0, 1]$.

(iii) $C(u_1, \ldots, u_d) = 0$ for all $(u_1, \ldots, u_d) \in [0, 1]^d$ with $u_i = 0$ for some $i$.

(iv) Given $a_1, \ldots, a_d \in [0, 1]$ and $b_1, \ldots, b_d \in [0, 1]$ such that $a_i \leq b_i$ for all $i \in \{1, \ldots, d\}$, we have

$$\mathbb{P}(U_1 \in [a_1, b_1], \ldots, U_d \in [a_d, b_d]) = \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1,i_1}, \ldots, u_{d,i_d}) \geq 0,$$

where $u_{j,1} = a_j$ and $u_{j,2} = b_j$ for all $j \in \{1, \ldots, d\}$.

Furthermore, any copula $C$ is Lipschitz-continuous and uniformly equicontinuous on its domain as shown in Theorem $3.2.4$ from Nelsen (2006). Thus the class of copulas is compact in $\mathbb{R}^d$.

Another important fact about copulas is a set of bounds due to Maurice Fréchet (1878–1973) and Wassily Hoeffding (1914–91). These bounds serve to represent the most extreme types of dependence, namely comonotonicity and countermonotonicty.

**Theorem 1.1.1** (Fréchet–Hoeffding bounds). *Let $C : [0,1]^d \to [0,1]$ be a copula. Then for all $u_1, \ldots, u_d \in [0,1]$, we have the following bounds:*

$$\max\left(1 - d + \sum_{i=1}^{d} u_i, 0\right) \leq C(u_1, \ldots, u_d) \leq \min(u_1, \ldots, u_d).$$

*Proof.* Let $U = (U_1, \ldots, U_d)$ be a random variable with standard uniform margins and copula $C$. Then, for any $u_1, \ldots, u_d \in [0,1]$,

$$C(u_1, \ldots, u_d) = \mathbb{P}(U_1 \leq u_1, \ldots, U_d \leq u_d) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{d} \{U_i > u_i\}\right) \geq 1 - \sum_{i=1}^{d} \mathbb{P}(U_i > u_i).$$

Given that $\mathbb{P}(U_i > u_i) = 1 - u_i$ for each $i \in \{1, \ldots, d\}$, we get

$$C(u_1, \ldots, u_d) \geq 1 - d + \sum_{i=1}^{d} u_i.$$

Moreover, for each $i \in \{1, \ldots, d\}$, one has

$$\mathbb{P}(U_1 \leq u_1, \ldots, U_d \leq u_d) \leq \mathbb{P}(U_i \leq u_i) = u_i,$$

and hence

$$C(u_1, \ldots, u_d) = \mathbb{P}(U_1 \leq u_1, \ldots, U_d \leq u_d) \leq \min(u_1, \ldots, x_d).$$

This completes the argument. □

As previously mentioned, a seminal result by Sklar (1959) explains why copulas play a key role in dependence modeling and how they can be used to characterize and describe the dependence relation between a collection of continuous random variables.

**Theorem 1.1.2** (Sklar). *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a continuous $d$-dimensional random vector and let $F_1, \ldots, F_d$ be the marginal cumulative distribution functions of $X_1, \ldots, X_d$, respectively. Then there exists a unique copula $C : [0,1]^d \to [0,1]$ such that, for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we*

*have*

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}. \tag{1.1}$$

*Conversely, if $C$ is a $d$-dimensional copula and $F_1, \ldots, F_d$ are univariate distribution functions, then $F$ as defined above is a joint distribution function with marginal distributions $F_1, \ldots, F_d$.*

A similar result also exists for survival functions. Namely if $\bar{F}$ is a $d$-variate continuous survival function with marginal survival functions $\bar{F}_1, \ldots, \bar{F}_d$, then there exists a unique copula $\bar{C}$, termed the survival copula, such that, for all $x_1, \ldots, x_d \in \mathbb{R}$,

$$\bar{F}(x_1, \ldots, x_d) = \bar{C}\{\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)\}.$$

In the case in which the random variables $X_1, \ldots, X_d$ are discontinuous, a similar decomposition holds, but the copula in (1.1) is not unique; see Genest & Nešlehová (2007).

Thanks to Sklar's decomposition theorem, when working with a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, the copula of $\boldsymbol{X}$ characterizes the dependence relation between the margins of $\boldsymbol{X}$. This theorem also has important implications for modeling: given a collection of variables $X_1, \ldots, X_d$, one can model the dependence between these variables by specifying their marginal distributions and a copula $C$ to define the joint distribution $F$ of $(X_1, \ldots, X_d)$ with formula (1.1) from Sklar's theorem.

The copula of a random vector, as specified by Sklar's theorem, is invariant under strictly monotonic transformations of the margins of the vector. Indeed, for any strictly increasing transformations $T_1, \ldots, T_d$, the copulas of $(X_1, \ldots, X_d)$ and $(T_1(X_1), \ldots, T_d(X_d))$ are the same. For a proof of the bivariate case of this result, which is identical to the proof for $d$-variate vectors, see Theorem 2.4.3 from Nelsen (2006). This property is particularly useful to sample from multivariate random variables with a particular copula, as it allows us to sample from a joint distribution with the desired copula and with standard uniform margins and then transform the margins using probability integral transformations to assign the desired marginal distributions to the sample. Since cumulative distribution

functions (cdfs) are non-decreasing, this transformation of the margins of the standard uniform vector does not affect the dependence structure.

## 1.2 Examples of copulas

There is a wide variety of copulas which represent different types of dependence relations. For instance, the simplest copula is the independence copula, defined for all $u_1, \ldots, u_d \in [0,1]$, by

$$C_I(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i.$$

It characterizes the mutual independence within a collection $X_1, \ldots, X_d$ of continuous variables. That is, the components of the vector $(X_1, \ldots, X_d)$ are mutually independent if and only if the copula of $(X_1, \ldots, X_d)$ is the independence copula.

In the light of Theorem 1.1.1, another pair of important functions are the upper and lower Fréchet–Hoeffding bounds, respectively defined, for all $u_1, \ldots, u_d \in [0,1]$, by

$$M_d(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d)$$

and

$$W_d(u_1, \ldots, u_d) = \max\{1 - d + u_1 + \cdots + u_d, 0\}.$$

Note that $M_d$ is always a copula but that $W_d$ is only a copula in dimension $d = 2$. These bounds are meant to represent two "extreme" dependence relations, comonotonicity and countermonotonicity. More specifically, continuous random variables $X_1, \ldots, X_d$ with distribution functions $F_1, \ldots, F_d$ are said to be comonotonic if and only if $M_d$ is the copula of the vector $(X_1, \ldots, X_d)$, in which case $F_1(X_1) = \cdots = F_d(X_d)$ almost surely. Similarly, the variables $X_1$ and $X_2$ are countermonotonic if and only if $W_2$ is the copula of the pair $(X_1, X_2)$, whence $F_2(X_2) = 1 - F_1(X_1)$ almost surely.

Copulas are often also classified into families that share similar construction and properties. Some of the most widely used copulas belong to the elliptical, Archimedean, and extreme-value copula families, which are briefly reviewed next.

### 1.2.1 Elliptical copulas

Elliptical copulas are referred to as such because they are derived from elliptical distributions. The two most common examples of elliptical distributions are the multivariate Gaussian $\mathcal{N}(\mu, \Sigma)$ and the multivariate Student $t$ distribution $t_v(\mu, \Sigma)$, both of which depend on a mean vector $\mu$ and a dispersion matrix $\Sigma$. However, all choices of $\mu$ lead to the same copula, so one can take $\mu = 0$ without loss of generality. Furthermore, owing to the invariance of copulas to monotone increasing transformations, the matrix $\Sigma$, which is symmetric and positive semi-definite by construction, can be taken to be a correlation matrix $R$ without loss of generality.

**Gaussian copulas**: Let $\Phi_R : \mathbb{R}^d \to [0, 1]$ be the joint distribution function of the multivariate Gaussian distribution with correlation matrix $R$ and mean vector $0$. Let $\Phi^{\leftarrow}$ be the quantile function of the univariate standard Gaussian distribution. Then the copula $C_R : [0, 1]^d \to [0, 1]$ defined, for all $u_1, \ldots, u_d \in (0, 1)$, by

$$C_R(u_1, \ldots, u_d) = \Phi_R\{\Phi^{\leftarrow}(u_1), \ldots, \Phi^{\leftarrow}(u_d)\}$$

is called the Gaussian copula with correlation matrix $R$.

**Student-$t$ copulas**: Let $T_{v,R} : \mathbb{R}^d \to [0, 1]$ be the joint distribution function of the centered multivariate Student $t$ distribution with correlation matrix $R$ and $v$ degrees of freedom. Let $t_v^{\leftarrow}$ be the quantile function of the univariate Student $t$ distribution, $t_v$, with $\nu$ degrees of freedom. Then the copula $C_{v,R} : [0, 1]^d \to [0, 1]$ defined, for all $u_1, \ldots, u_d \in (0, 1)$, by

$$C_{v,R}(u_1, \ldots, u_d) = T_{v,R}\{t_v^{\leftarrow}(u_1), \ldots, t_v^{\leftarrow}(u_d)\}$$

is called the Student $t$ copula with correlation matrix $R$ and $v$ degrees of freedom.

Every multivariate random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with $t_v$ distribution and correlation matrix $R$ has a $t_{v,R}$ copula, and every $\boldsymbol{X} = (X_1, \ldots, X_d)$ with Gaussian distribution and correlation matrix $R$ has a Gaussian copula with the corresponding correlation matrix. As the Normal and Student $t$ distributions have appealing properties and are widely used in modeling, the corresponding elliptical copulas are also some of the most commonly used in practice.

### 1.2.2 Archimedean copulas

Another class of copulas with widespread use in dependence modeling are Archimedean copulas. These copulas are characterized by a map $\psi : [0, 1] \to [0, \infty)$, called the generator function. A generator $\psi$ must be continuous, convex, strictly decreasing and such that $\psi(1) = 0$. Moreover, let $\psi^{[-1]}$ be the pseudo-inverse of $\psi$, where

$$\psi^{[-1]}(x) = \begin{cases} \psi^{-1}(x) & \text{if } 0 \leq x \leq \psi(0), \\ 0 & \text{if } \psi(0) \leq x \leq \infty. \end{cases}$$

A copula $C : [0, 1]^d \to [0, 1]$ is said to be Archimedean if there exists a generator $\psi : [0, 1] \to \mathbb{R}$ such that, for all $u_1, \ldots, u_d \in (0, 1)$,

$$C(u_1, \ldots, u_d) = \psi^{[-1]}\{\psi(u_1) + \cdots + \psi(u_d)\}. \tag{1.2}$$

The above conditions on the generator are necessary and sufficient to ensure that $C$ is a copula in dimension $d = 2$. In higher dimensions, however, more conditions are needed, which were characterized by McNeil & Nešlehová (2009).

Examples of Archimedean copulas are the Gumbel copula with parameter $\theta \in [1, \infty)$, whose generator is defined, for all $x \in (0, 1)$, by $\psi(x) = |\ln(x)|^\theta$. Upon substituting this

choice of $\psi$ into Equation (1.2), one finds that, for all $u_1, \ldots, u_d \in (0,1)$,

$$C_{Gu(\theta)}(u_1, \ldots, u_d) = \exp\left[ -\left\{ \sum_{i=1}^{d} |\ln(u_i)|^{\theta} \right\}^{1/\theta} \right].$$

Another example is the Clayton copula with parameter $\theta \in (0, \infty)$, whose generator is given, for all $x \in (0,1)$, by $\psi(x) = (x^{-\theta} - 1)/\theta$. Upon substituting this choice of $\psi$ into Equation (1.2), one finds that, for all $u_1, \ldots, u_d \in (0,1)$,

$$C_{Cl(\theta)}(u_1, \ldots, u_d) = \left\{ \sum_{i=1}^{d} (u_i^{-\theta} - 1) + 1 \right\}^{-1/\theta}.$$

In addition to Archimedean copulas, extreme-value copulas are a frequently used class of copulas that are of great interest in the field of extreme-value theory. We expand on extreme-value copulas in Chapter 5.

## 1.3 Copula-based dependence measures

As copulas can be used to characterize the dependence structure of collections of random variables, various measures of dependence between pairs of random variables can be expressed in terms of their underlying copula. In particular, two of the most popular measures of rank correlation, Kendall's $\tau$ and Spearman's $\rho$, depend entirely on the copula of the pair of continuous random variables being examined.

Given a pair $(X, Y)$ of continuous random variables, Kendall's $\tau$ is defined as

$$\tau(X, Y) = \mathbb{P}\{(X - X^*)(Y - Y^*) > 0\} - \mathbb{P}\{(X - X^*)(Y - Y^*) < 0\},$$

where $(X^*, Y^*)$ is an independent copy of the pair $(X, Y)$. Kendall's $\tau$ is a measure of ordinal association. It compares the probability that two random pairs will be concordant vs. the probability that they are discordant. In the case where both $X$ and $Y$ are continuous,

as shown in Theorem $5.1.1$ of Nelsen (2006), Kendall's $\tau$ can be written as

$$\tau(X,Y) = -1 + 4\int_0^1 \int_0^1 C(u,v)dC(u,v),$$

in terms of the unique copula $C$ of the pair $(X,Y)$.

Another commonly used rank correlation measure is Spearman's $\rho$. Given continuous random variables $X$ and $Y$ with cumulative distribution functions $F_X$ and $F_Y$, respectively, Spearman's $\rho$ is defined as

$$\rho(X,Y) = \rho_p\{F_X(X), F_Y(Y)\},$$

where $\rho_p$ stands for the standard Pearson correlation measure. In other words, Spearman's $\rho$ is the regular Pearson correlation of the probability integral transform of the pairs of variables. Thus, letting $U_X = F_X(X)$ and $U_Y = F_Y(Y)$, which are both uniformly distributed on the interval $(0,1)$, one has

$$\rho(X,Y) = \text{cor}(U_X,U_Y) = \frac{\text{cov}(U_X,U_Y)}{\sqrt{\text{var}(U_X)\text{var}(U_Y)}},$$

where

$$\text{cov}(U_X,U_Y) = \text{E}[\{U_X - \text{E}(U_X)\}\{U_Y - \text{E}(U_Y)\}] = \text{E}(U_XU_Y) - \frac{1}{4}$$

while

$$\text{var}(U_X) = \text{var}(U_Y) = \frac{1}{12},$$

which gives the following formula for Spearman's $\rho$ provided by Theorem $5.1.6$ of Nelsen (2006):

$$\rho(X,Y) = -3 + 12\int_0^1 \int_0^1 C(u,v)dvdu.$$

Thus both $\tau$ and $\rho$ are entirely defined by the unique copula of the pair $(X,Y)$.

Similarly as Pearson's correlation measure, both $\tau$ and $\rho$ take values in $[-1, 1]$. If $X$ and $Y$ are independent, then $\tau(X, Y) = \rho(X, Y) = 0$; however, the reverse implication does not hold.

These measures are also invariant to monotonic transformations. Indeed, for any monotonic transformation $T : \mathbb{R} \to \mathbb{R}$, the copulas of $(X, Y)$ and $(T(X), T(Y))$ are the same. Then $\tau\{T(X), T(Y)\} = \tau(X, Y)$ and $\rho\{T(X), T(Y)\} = \rho(X, Y)$ because $\tau$ and $\rho$ are entirely determined by the common, unique copula of $(X, Y)$ and $(T(X), T(Y))$.

In particular, it $Y$ and $X$ are comonotonic, i.e., $Y = T(X)$ for some strictly monotonic function $T$, then $T$ is strictly increasing if and only if $\tau(X, Y) = \rho(X, Y) = 1$ and $T$ is strictly decreasing if and only if $\tau(X, Y) = \rho(X, Y) = -1$.

### 1.3.1 Tail dependence measures

In certain applications involving extreme values, it is relevant to study the degree of association between simultaneously large, or low, values of two dependent random variables. These coefficients describe the amount of dependence in the upper-right-quadrant or lower-left-quadrant of a bivariate distribution. There are two kinds of tail dependence measures, which are defined as follows, given two continuous random variables $X$ and $Y$ with cumulative distribution functions $F_X$ and $F_Y$, respectively.

(i) **Upper tail dependence**: The coefficient of upper tail dependence for the pair $(X, Y)$ is given by the following limit, provided it exists:

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

where $C$ is the unique copula of the pair $(X, Y)$.

(ii) **Lower tail dependence**: Similarly, the coefficient of lower tail dependence for the pair $(X, Y)$ is given by the following limit, provided it exists:

$$\lambda_L(X,Y) = \lim_{x \searrow 0} \mathbb{P}(X \leq F_X^{\leftarrow}(x) \mid Y \leq F_Y^{\leftarrow}(x)\} = \lim_{u \searrow 0} \frac{C(u,u)}{u},$$

where $C$ is the unique copula of the pair $(X, Y)$. It is clear once again that the copula of $(X, Y)$ entirely determines its lower and upper tail dependence coefficients.

## 1.4 Copula estimators

Suppose that a random sample of size $n$ from $\boldsymbol{X} = (X_1, \ldots, X_d)$ is available. For each $i \in \{1, \ldots, n\}$, let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$ be the $i$th observation in this sample. Assuming that the margins $F_1, \ldots, F_d$ of $X_1, \ldots, X_d$ are absolutely continuous with densities $f_1, \ldots, f_d$, respectively, there are various methods to estimate the copula associated with $\boldsymbol{X}$. We present three such approaches using parametric, semi-parametric, and nonparametric rank-based methods.

### 1.4.1 Parametric approach

Assume that there exists a copula $C_\theta$ for $\boldsymbol{X}$ which belongs to a parametric family of copulas indexed by a parameter $\theta$ which could be real-valued, vector-valued, or even matrix-valued as in the case of Gaussian copulas. Assume that the copula $C_\theta$ is absolutely continuous and let $c_\theta$ be the corresponding density. Given estimates for the parameters of the marginal distributions $\theta_1, \ldots, \theta_d$, the dependence parameter $\theta$ can be estimated using the method of maximum likelihood. By Sklar's theorem and the chain rule, upon differentiating the distribution function $F$ of $\boldsymbol{X}$, we obtain a formula for the corresponding density $f$ given, for all $x_1, \ldots, x_d \in \mathbb{R}$, by

$$f(x_1, \ldots, x_d) = c_\theta\{F_1(x_1), \ldots, F_d(x_d)\} \prod_{j=1}^{d} f_j(x_j).$$

Therefore, given observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})$ for each $i \in \{1, \ldots, n\}$, we may use this expression to obtain a formula for the log-likelihood $L$ for $\theta$ and $\theta_1, \ldots, \theta_d$, viz.

$$L(\theta, \theta_1, \ldots, \theta_d) = \sum_{i=1}^{n} \left[ \ln[c_\theta\{F_1(x_{i1} \mid \theta_1), \ldots, F_d(x_{id} \mid \theta_d)\}] + \sum_{j=1}^{d} \ln\{f(x_{ij} \mid \theta_j)\} \right]. \quad (1.3)$$

Estimators for $\theta$, $\theta_1, \ldots, \theta_d$ can be obtained by maximizing expression (1.3). This approach has the main disadvantage that the process can be computationally expensive when $d$ is large, and that the estimate of $\theta$ can be biased if the marginal distributions are misspecified.

### 1.4.2 Semi-parametric approach

The semi-parametric approach is an adaptation of the MLE method to account for some of its drawbacks. Instead of using parametric models for $F_1, \ldots, F_d$, the marginal distributions of $X_1, \ldots, X_d$ are estimated using empirical distributions.

Let $(X_{11}, \ldots, X_{1d}), \ldots, (X_{n1}, \ldots, X_{nd})$ be a random sample of size $n$ from the distribution of the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$. Then for each $j \in \{1, \ldots, d\}$, we use empirical cumulative distribution functions to estimate the marginal distributions of each $X_j$. The latter is given, for each $x \in \mathbb{R}$, by

$$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_{ij} \leq x),$$

where $\mathbb{I}(X_{ij} \leq x)$ is an indicator function for the event $\{X_{ij} \leq x\}$. Note that for any $i \in \{1, \ldots, d\}$ and $\ell \in \{1, \ldots, n\}$, one has

$$\hat{F}_j\big(X_{\ell j}\big) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\big(X_{ij} \leq X_{\ell j}\big) = \frac{1}{n} R_{\ell j},$$

where $R_{\ell j}$ is the rank of $X_{\ell j}$ among $X_{1j}, \ldots, X_{nj}$. Next, we replace

$$\sum_{i=1}^{n} \ln \left[ c_\theta \{ F_1(x_{i1} \mid \theta_1), \ldots, F_d(x_{id} \mid \theta_d) \} \right]$$

by

$$\sum_{j=1}^{n} \ln \left[ c_\theta \{ \hat{F}_1(x_{i1}), \ldots, \hat{F}_d(x_{id}) \} \right]$$

in (1.3). The replacement of parametric models by empirical estimators allows us to avoid the bias caused by the misspecification of the marginal distributions.

### 1.4.3 Nonparametric approach

Given a random sample from a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, its unknown but unique copula $C$ can be approximated using a rank-based estimator called the empirical copula. Given that $C$ is the distribution function of $(F_1(X_1), \ldots, F_d(X_d))$, we may estimate $C$ using empirical probability measures.

We use the following estimator for $C$, known as the empirical copula. It is defined, at any $u_1, \ldots, u_d \in (0, 1)$, by

$$\hat{C}_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\big\{ \hat{F}_1\big(X_{i1}\big) \leq u_1, \ldots, \hat{F}_d\big(X_{id}\big) \leq u_d \big\}.$$

Note that the empirical copula estimator assigns a weight of $1/n$ to each atom of the form $(R_{1j}/n, \ldots, R_{dj}/n)$, where for each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, d\}$, $R_{ij}$ is the rank of $X_{ij}$ among $X_{1j}, \ldots, X_{nj}$. Thus, the empirical copula estimator is entirely defined by the component-wise ranks of the points in the sample.

Note that given that $\hat{C}_n$ is discontinuous, it is not a copula per se. Nevertheless, empirical copulas possess various interesting asymptotic properties and can be used to develop inference procedures, e.g., to compare copulas via the test underlined in Rémillard & Scaillet (2007) or to test certain properties of the underlying copula of a random vector such as the test of asymmetry presented in Genest & Nešlehová (2013).

**Definition** (Empirical copula process)**.** For a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with unique underlying copula $C$, let $(X_{11}, \ldots, X_{1d}), \ldots, (X_{n1}, \ldots, X_{nd})$ be a random sample from $\boldsymbol{X}$ and let $C_n$ be the empirical copula estimator for $C$ based on this sample. Then the stochastic process $\mathbb{C}_n$ defined, for all $u_1, \ldots, u_d \in [0, 1]$, by

$$\mathbb{C}_n(u_1, \ldots, u_d) = \sqrt{n}\,\{C_n(u_1, \ldots, u_d) - C(u_1, \ldots, u_d)\}$$

is called the empirical copula process for $\boldsymbol{X}$.

Similarly to empirical distribution functions, the empirical copula process has very useful asymptotic properties. Under the regularity conditions for the underlying copula $C$ from Proposition $3.1$ in Segers (2012), the empirical copula process $\mathbb{C}_n$ converges weakly to a centered Gaussian process.

For each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, d\}$, let $U_{ij} = F_j(X_{ij})$. Further, let $G_n$ denote the empirical distribution function of the unobserved random sample

$$U_1 = (U_{11}, \ldots, U_{1d}), \ldots, U_n = (U_{n1}, \ldots, U_{nd}). \tag{1.4}$$

Next, consider the process defined for, all $u_1, \ldots, u_d \in [0, 1]$, by

$$\alpha_n(u_1, \ldots, u_d) = \sqrt{n}\,\{G_n(u_1, \ldots, u_d) - C(u_1, \ldots, u_d)\}.$$

Then, there exists a centered, tight, Gaussian process $\alpha$ on $[0, 1]^d$ such that $\alpha_n \rightsquigarrow \alpha$ as $n \to \infty$ on $L^\infty[0, 1]^d$, where $L^\infty[0, 1]^d$ is the space of bounded functions on $[0, 1]^d$ equipped with the supremum norm. Moreover, the covariance function of $\alpha$ is given, for all $u_1, \ldots, u_d$, $v_1, \ldots, v_d \in [0, 1]$, by

$$\begin{aligned}\operatorname{cov}\{&\alpha(u_1, \ldots, u_d), \alpha(v_1, \ldots, v_d)\} \\ &= C\{\min(u_1, v_1), \ldots, \min(u_d, v_d)\} - C(u_1, \ldots, u_d)C(v_1, \ldots, v_d).\end{aligned}$$

The stochastic process $\alpha$ so defined is called a $C$-Brownian sheet.

For each $j \in \{1, \ldots, d\}$, let $\dot{C}_j$ be the partial derivative of $C$ with respect to its $j$th component. Under the assumption that, for each $j \in \{1, \ldots, d\}$, $\dot{C}_j$ exists and is continuous on the set $V_{d,j} = \{(u_1, \ldots, u_d) \in [0,1]^d : 0 < u_j < 1\}$, this weak convergence result can be generalized as follows to distributions with arbitrary continuous margins.

**Theorem 1.4.1.** *Let $\mathbb{C}_n = \sqrt{n}\,(C_n - C)$ be the empirical copula process on $[0,1]^d$. Under the above assumptions about the continuity of the partial derivatives of $C$, $\mathbb{C}_n$ converges weakly, as $n \to \infty$, to a centered Gaussian process $\mathbb{C}$ on $[0,1]^d$ given, for all $u_1, \ldots, u_d \in [0,1]$, by*

$$\mathbb{C}(u_1, \ldots, u_d) = \alpha(u_1, \ldots, u_d) - \sum_{j=1}^{d} \dot{C}_j(u_j)\alpha(1, \ldots, u_j, \ldots, 1),$$

*where $\alpha$ is a $C$-Brownian sheet.*

The asymptotic properties of empirical copula processes are used for a variety of inference procedures about the copula of a random sample.

One intriguing fact about the empirical copula process is that under certain conditions, it can have a smaller asymptotic covariance function than the standard empirical process based on observations (1.4) from the copula, thereby leading to more precise inference. This phenomenon was studied by Genest & Segers (2010) in the bivariate case, and by Genest et al. (2019) in higher dimensions.

# Chapter 2

# Aggregation-tree copula models and conditional independence assumptions

Copula models are challenging in the high-dimensional setting. Indeed, in high dimensions, copulas tend to have complicated densities, traditional models such as elliptical and Archimedean copulas lack flexibility, and the inference procedures to validate these models can also be quite complex. Several approaches have been proposed to remedy the limitations of such models in high dimensions. Most notably, vine copula constructions are discussed in the books by Kurowicka & Joe (2011) and Czado (2019); see also the book by Joe (2014) and the review paper by Czado & Nagler (2021).

In this chapter, we present an alternative approach, which we refer to as aggregation-tree copula models, originally proposed by Arbenz et al. (2012) and expanded upon by Côté & Genest (2015). We examine the necessary conditional independence assumptions which ensure that an aggregation-tree model generates a unique and well-defined multivariate distribution.

## 2.1 Aggregation-tree copula models

Aggregation-tree copula models are a simple and attractive alternative to other solutions to the challenge of copula models for high-dimensional data. Indeed, the previously mentioned approaches suffer from either being limited to a restrictive range of dependence structures which they can model, or they require complex inference procedures. As will be seen, aggregation-tree models are more flexible than traditional multivariate copula families. They can accommodate a much larger range of dependence structures and they are quite simple to implement even in high-dimensional settings.

Aggregation-tree models revolve around the use of low-dimensional copulas to define recursively the dependence structure of aggregated versions of the original variables until the dependence structure of the entire collection of variables has been specified. We will focus on the version of these models that relies on bivariate copulas exclusively, similar to what was presented by Côté & Genest (2015).

Given a collection $X_1, \ldots, X_d$ of continuous random variables, the construction of an aggregation-tree copula model for the random vector $(X_1, \ldots, X_d)$ requires four elements:

(i) an aggregation function $a$;

(ii) a tree structure $\mathcal{T}$ for the model;

(iii) a collection of marginal distributions $F_1, \ldots, F_d$, and

(iv) a collection $C_1, \ldots, C_{d-1}$ of bivariate copulas.

These are described in turn.

### Aggregation functions

The choice of aggregation function is a crucial part of the model construction process. An aggregation function is a map $a : \mathbb{R}^2 \to \mathbb{R}$ that allows us to combine multiple random variables into one. This choice should be ideally determined by the context of the data

and what suits the researcher. However, there are certain properties that are required for an aggregation function to be suitable for these types of models:

(i) Commutativity: $a(x, y) = a(y, x)$ for all $x, y \in \mathbb{R}$.

(ii) Transitivity: $a\{a(x, y), z\} = a\{x, a(y, z)\}$ for all $z \in \mathbb{R}$.

(iii) Homogeneity: $a(cx, cy) = ca(x, y)$ for any $c \in \mathbb{R}$.

Another desirable property is that $a(x, x) = x$. These properties ensure that the order in which the variables are aggregated does not matter. In particular, therefore, a unique meaning is ascribed to the overall aggregate $a(x_1, \ldots, x_k)$, as any successive pair-wise aggregation of $x_1, \ldots, x_k$ will lead to the same result. Examples of aggregation functions that satisfy these properties are $a(x, y) = x + y$, $a(x, y) = xy$, $a(X, Y) = \min(x, y)$, $a(x, y) = \max(x, y)$, as well as means and quasi-arithmetic means.

Ideally, the aggregation function should be chosen such that the combinations of variables created from the aggregation process have a meaningful interpretation in the context of the analysis. For instance, Arbenz et al. (2012) and Côté & Genest (2015) used $a(x, y) = x + y$ as the aggregation function in their papers. Both articles were presented in the context of risk modeling, where the random variables $X_1, \ldots, X_d$ represented individual risks. Accordingly,

$$a(X_{i_1}, \ldots, X_{i_k}) = X_{i_1} + \cdots + X_{i_k}$$

was a combined risk. In the context of climatology, however, if we were working with variables that represent the level of precipitation at a hydrological station across a certain time-frame, for example, then using $a(x, y) = \max(x, y)$ might be a more sensible choice given that the aggregated variables have a useful interpretation as the maximum level of precipitation in a specific region.

**Tree structure**

The tree structure is the other element of these models that describes the aggregation process. The tree structure $\mathcal{T}$ represents the way in which the variables $X_1, \ldots, X_d$ are aggregated. It consists of a sequence of collections of index sets $\mathcal{T}_1, \ldots, \mathcal{T}_d$, which represent the order in which the variables are aggregated at each step of the aggregation process. We always have $\mathcal{T}_1 = \{A_{11}, \ldots, A_{d1}\}$ with $A_{j1} = \{j\}$ for all $j \in \{1, \ldots, d\}$, and $\mathcal{T}_d = \{\{1, \ldots, d\}\}$. The construction of an aggregation-tree model which uses bivariate copulas exclusively proceeds as follows:

For each $i \in \{1, \ldots, d-1\}$:

(1) Select a pair of indices $\ell_i \neq k_i \in \{1, \ldots, d+1-i\}$.

(2) The index sets $A_{\ell_i i}$ and $A_{k_i i}$ are combined, as we set $A_{1(i+1)} = A_{\ell_i i} \bigcup A_{k_i i}$.

(3) The remaining index sets $A_{ji}$, where $j \notin \{\ell_i, k_i\}$ are relabeled as $A_{2(i+1)}, \ldots, A_{(d-i)(i+1)}$ and we set $\mathcal{T}_{i+1} = \{A_{1(i+1)}, \ldots, A_{(d-i)(i+1)}\}$.

(4) Choose a copula $C_{A_{j_i i} A_{k_i i}}$ to model $(M_{A_{ji}}, M_{A_{ki}})$, where $M_A = a(X_\ell : \ell \in A)$.

This process can be modified by combining a different number of index sets in Step $(2)$ and using copulas of the corresponding dimension of the aggregated variables in Step $(4)$. We refer to these models as aggregation-tree models given that the tree structure can be represented as a tree with a single root, $d$ leaf nodes, and $d-1$ branching nodes with exactly two children nodes. In this tree representation, the leaf nodes represent the marginal random variables and the branching nodes represent the aggregated random variables. The copulas characterize the dependence structure of each of the aggregated variables associated with the children nodes which are combined to create a branching node.

The following example serves to illustrate how the tree structure defines the dependence structure in the model.

**Example 2.1.1.** Consider an aggregation-tree model for a $4$-dimensional random vector $(X_1, X_2, X_3, X_4)$ with aggregation function $a(x, y) = x + y$ and tree structure given by $\mathcal{T}_1 =$

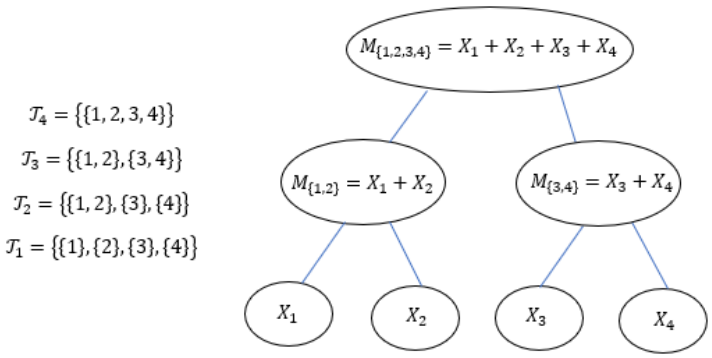


**Figure 2.1:** Example of a tree structure for four variables using partial sums as the aggregation function in Example 2.1.1.

$\{\{1\},\{2\},\{3\},\{4\}\}$, $\mathcal{T}_2 = \{\{1,2\},\{3\},\{4\}\}$, $\mathcal{T}_3 = \{\{1,2\},\{3,4\}\}$ and $\mathcal{T}_4 = \{\{1,2,3,4\}\}$. To fit an aggregation-tree model with this tree structure and aggregation function, we require a copula $C_{\{1\}\{2\}}$ and marginal distributions $F_1$ and $F_2$ to model $(X_1, X_2)$, a copula $C_{\{3\}\{4\}}$ and marginal distributions $F_3$ and $F_4$ to model $(X_3, X_4)$ and finally, a copula $C_{\{1,2\}\{3,4\}}$ to model $(X_1 + X_2, X_3 + X_4)$.

As argued by Arbenz et al. (2012) and Côté & Genest (2015), aggregation-tree models require a crucial assumption about the conditional independence of the node variables in the tree structure. Such an assumption is needed to ensure that we obtain a unique and well-defined joint distribution from the model. In the following sections, we explore different formulations of the conditional independence assumption and provide proofs that they are sufficient to obtain a unique and well-defined distribution. Before we proceed, we introduce some notation.

### Notation

To facilitate our discussion of properties and proofs related to aggregation-tree copula models, we introduce some notation. For each $A \subseteq D = \{1, \ldots, d\}$, let $\mathbf{F}_A$ denote the joint distribution function of the vector $\boldsymbol{X}_A = (X_i : i \in A)$ and let $F_A$ denote the distribution function of $M_A = a(X_i : i \in A)$. Moreover, for any $A \subset D$, let $\bar{A} = D \setminus A$.

We use $A \in \mathcal{T}$ to denote $A \in \bigcup_{i=1}^{d} \mathcal{T}_i$. We denote the index sets for the branching nodes of the tree structure (aggregated variables) by $\mathcal{B}(\mathcal{T}) = \{A \in \mathcal{T} : |A| > 1\}$, and for the leaf nodes (marginal variables) by $\mathcal{L}(\mathcal{T}) = \{A \in \mathcal{T} : |A| = 1\}$.

For any aggregated variable $M_A$, $A \in \mathcal{B}(\mathcal{T})$, we denote the index sets of the descendants of $M_A$ by $\mathcal{D}(A) = \{B \in \mathcal{T} : B \subset A\}$ and the index sets of the leaf descendants of $M_A$ by $\mathcal{L}(A) = \{B \in \mathcal{L}(\mathcal{T}) : B \subset A\}$. If a branching node $M_A$ is obtained by combining the nodes $M_{B_1}$ and $M_{B_2}$, we call $M_{B_1}$ and $M_{B_2}$ children nodes of $M_A$ and we use $\mathcal{C}(A)$ to denote the index sets of the children of $M_A$.

## 2.2 Conditional expectation: Preliminaries

Before presenting the conditional independence assumption for aggregation-tree models, we introduce some preliminary results about conditional expectation and conditional independence which are used in proofs. First, we define conditional expectations in terms of $\sigma$-fields on the underlying probability space.

**Definition 2.2.1** (Conditional expectation)**.** Let $(\Omega, \mathcal{S}, \mathbb{P})$ be a probability space and let $m(\mathcal{S})$ be the collection of real-valued random variables on this probability space. Let $X \in L^1(\mathbb{P}) = \{X \in m(\mathcal{S}) : \mathrm{E}(|X|) < \infty\}$ and let $\mathcal{A}$ be a sub-field of $\mathcal{S}$. There exists an almost surely unique linear operator $\mathrm{E}(\cdot \mid \mathcal{A}) : L^1(\mathbb{P}) \to L^1(\mathbb{P})$ called the conditional expectation (conditioning on $\mathcal{A}$) such that, for all $A \in \mathcal{A}$,

$$\int_A \mathrm{E}(X \mid \mathcal{A}) d\mathbb{P} = \int_A X d\mathbb{P}.$$

Similarly, given $Y \in L^1(\mathbb{P})$, let $\sigma(Y)$ denotes the $\sigma$-field generated by $Y$, i.e., the smallest $\sigma$-field containing $\Sigma = \{Y^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel $\sigma$-field on $\mathbb{R}$. We define $\mathrm{E}(\cdot \mid Y) : L^1(\mathbb{P}) \to L^1(\mathbb{P})$ by $\mathrm{E}(X \mid Y) = \mathrm{E}\{X \mid \sigma(Y)\}$. For $Y_1, \ldots, Y_n \in L^1(\mathbb{P})$, we define $\mathrm{E}(\cdot \mid Y_1, \ldots, Y_n) : L^1(\mathbb{P}) \to L^1(\mathbb{P})$ with $\mathrm{E}(X \mid Y_1, \ldots, Y_n) = \mathrm{E}\{X \mid \bigvee_{i=1}^{n} \sigma(Y_i)\}$,

where

$$\bigvee_{i=1}^{n} \sigma(Y_i) = \sigma\left\{\bigcup_{i=1}^{n} \sigma(Y_i)\right\} = \sigma(Y_1, \ldots, Y_n).$$

With this definition of conditional expectations in terms of $\sigma$-fields, conditional expectation possesses many important properties similar to those of regular expectation that we will make use of.

**Properties of conditional expectation**

The following facts are listed on p. 88 of the book by Williams (1991).

(i) Positivity: if $X \geq 0$, then $\mathrm{E}(X \mid \mathcal{A}) \geq 0$ almost surely.

(ii) $L^p$-contractivity: $\|\mathrm{E}(X \mid \mathcal{A})\|_p \leq \|X\|_p$ for any real $p \in [1, \infty)$, where $\|\cdot\|_p = \mathrm{E}(|\cdot|^p)^{1/p}$ is the standard $L^p(\mathbb{P})$ norm.

(iii) Monotone convergence: Let $\{X_i : i \in \mathbb{N}\}$ be any sequence of non-negative random variables such that $X_n$ converges monotonically to $X$ as $n \to \infty$ almost surely, i.e., $X_n \leq X_{n+1}$ for all $n \in \mathbb{N}$ and $X_n \to X$ as $n \to \infty$ (denoted as $X_n \nearrow X$). Then $\mathrm{E}(X_n \mid \mathcal{A})$ converges monotonically to $\mathrm{E}(X \mid \mathcal{A})$ almost surely as $n \to \infty$.

(iv) Dominated convergence: Let $\{X_i : i \in \mathbb{N}\}$ be any sequence of random variables such that $|X_n| \leq Z$ almost surely for all $n \in \mathbb{N}$ and for some $Z \in L^1(\mathbb{P})$. If $X_n \to X$ almost surely as $n \to \infty$, then $\mathrm{E}(X_n \mid \mathcal{A}) \to \mathrm{E}(X \mid \mathcal{A})$ almost surely as $n \to \infty$.

Well-known inequalities for expected values of random variables such as Jensen's and Markov's inequality also have equivalent versions for conditional expectations. There are other important properties for conditional expectation relating to the $\sigma$-field used for conditioning, as listed below:

(v) If $Y$ is measurable with respect to a $\sigma$-field $\mathcal{A}$, which we denote as $Y \in m(\mathcal{A})$, then $\mathrm{E}(YX \mid \mathcal{A}) = Y\mathrm{E}(X \mid \mathcal{A})$ for any $X$. More generally, for any measurable function $g$ with respect to $\mathcal{S}$, $\mathrm{E}\{g(Y)X \mid Y\} = g(Y)\mathrm{E}(X \mid Y)$. Moreover, if $X \in m(\mathcal{A})$, then $\mathrm{E}(X \mid \mathcal{A}) = X$ almost surely.

(vi) Tower property: Given any $\sigma$-fields $\mathcal{G}$ and $\mathcal{H}$ such that $\mathcal{G} \subset \mathcal{H} \subset \mathcal{S}$, one has

$$\mathrm{E}\big\{\mathrm{E}(X \mid \mathcal{H}) \mid \mathcal{G}\big\} = \mathrm{E}(X \mid \mathcal{G}) = \mathrm{E}\big\{\mathrm{E}(X \mid \mathcal{G}) \mid \mathcal{H}\big\}$$

almost surely.

The tower property is particularly useful and intuitive. It asserts that conditioning on a larger $\sigma$-field does not affect the expected value of a conditional expectation.

**Definition 2.2.2** (Conditional probability)**.** Let $\mathcal{F}$ be a sub-field of $\mathcal{S}$. Then the conditional probability $\mathbb{P}(A \mid \mathcal{F})$ is given, for any $A \in \mathcal{S}$, by

$$\mathbb{P}(A \mid \mathcal{F}) = \mathrm{E}(\mathbb{I}_A \mid \mathcal{F}).$$

In particular, for any collection $X, Y_1, \ldots, Y_n$ of random variables in $(\Omega, \mathcal{S}, \mathbb{P})$, and for any $A \in \mathcal{B}(\mathbb{R}^d)$, one has

$$\mathbb{P}(X \in A \mid Y_1, \ldots, Y_n) = \mathrm{E}\{\mathbb{I}_A(X) \mid Y_1, \ldots, Y_n\}.$$

Using this definition for conditional probabilities, we define conditional independence among random variables.

**Definition 2.2.3** (Conditional independence)**.** Let $\mathcal{F}$, $\mathcal{H}$ and $\mathcal{G}$ be sub-fields of $\mathcal{S}$. Then $\mathcal{F}$ and $\mathcal{H}$ are conditionally independent given $\mathcal{G}$, denoted $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$, if for any $A \in \mathcal{F}$ and $B \in \mathcal{H}$, the identity

$$\mathbb{P}(A \cap B \mid \mathcal{G}) = \mathbb{P}(A \mid \mathcal{G})\mathbb{P}(B \mid \mathcal{G})$$

holds almost surely.

Additionally, let $X$ and $Y$ be random variables on $(\Omega, \mathcal{S}, \mathbb{P})$. Then $X$ and $Y$ are conditionally independent given $\mathcal{G}$, denoted $X \perp\!\!\!\perp_{\mathcal{G}} Y$, if the $\sigma$-fields $\sigma(X)$ and $\sigma(Y)$ are conditionally independent given $\mathcal{G}$. Similarly, for any collection of random variables $Z_1, \ldots, Z_n$ on $(\Omega, \mathcal{S}, \mathbb{P})$, the random variables $X$ and $Y$ are conditionally independent

given $Z_1, \ldots, Z_n$, denoted $X \perp\!\!\!\perp_{Z_1,\ldots,Z_n} Y$, if the $\sigma$-fields $\sigma(X)$ and $\sigma(Y)$ are conditionally independent given $\bigvee_{i=1}^{n} \sigma(Z_i)$ .

From this definition, it follows directly that for any $\sigma$-fields $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{W}$ such that $\mathcal{G} \subset \mathcal{F}$, if we have $\mathcal{F} \perp\!\!\!\perp_{\mathcal{W}} \mathcal{H}$, then it follows that $\mathcal{G} \perp\!\!\!\perp_{\mathcal{W}} \mathcal{H}$. In particular, for a measurable function of random variables $X_1, \ldots, X_n$, let $W = g(X_1, \ldots, X_n)$. If $X_1, \ldots, X_n \perp\!\!\!\perp_{\mathcal{G}} Y$, then we get $W \perp\!\!\!\perp_{\mathcal{G}} Y$ because $\sigma(W) \subset \bigvee_{i=1}^{n} \sigma(X_i)$.

However, verifying the conditional independence between $\sigma$-fields or variables may be difficult from the definition alone. Theorem 8.9 of Kallenberg (2022), which is due to Joseph Doob (1910–2004), provides a very useful alternative characterization of conditional independence in terms of the ability to "drop" certain $\sigma$-fields from the conditioning. This result is stated and proved below for completeness, following the arguments from Kallenberg (2022) since similar techniques will be used in other proofs.

**Theorem 2.2.1** (Doob). *For any $\sigma$-fields $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$, the following statements are equivalent:*

*(i)* $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$;

*(ii)* $\mathbb{P}(\cdot \mid \mathcal{F} \vee \mathcal{G}) = \mathbb{P}(\cdot \mid \mathcal{G})$ *almost surely on* $\mathcal{H}$.

*Proof.* First, we prove that *(ii)* implies *(i)*. Assume $\mathbb{P}(H \mid \mathcal{F} \vee \mathcal{G}) = \mathbb{P}(H \mid \mathcal{G})$ almost surely for all $H \in \mathcal{H}$. Let $F \in \mathcal{F}$ and $H \in \mathcal{H}$. Then, by the tower property, one finds

$$\mathbb{P}(H \cap F \mid \mathcal{G}) = \mathrm{E}(\mathbb{I}_{H\cap F} \mid \mathcal{G}) = \mathrm{E}\{\mathrm{E}(\mathbb{I}_{H\cap F} \mid \mathcal{G} \vee \mathcal{F}) \mid \mathcal{G}\} = \mathrm{E}\{\mathbb{I}_F \mathrm{E}(\mathbb{I}_H \mid \mathcal{G} \vee \mathcal{F}) \mid \mathcal{G}\},$$

given that $\mathbb{I}_F \in m(\mathcal{G} \vee \mathcal{F})$. Moreover, using *(ii)*, $\mathrm{E}(\mathbb{I}_H \mid \mathcal{G} \vee \mathcal{F}) = \mathrm{E}(\mathbb{I}_H \mid \mathcal{G})$ almost surely. Hence, we have almost surely that

$$\mathrm{E}\{\mathbb{I}_F \mathrm{E}(\mathbb{I}_H \mid \mathcal{G} \vee \mathcal{F}) \mid \mathcal{G}\} = \mathrm{E}(\mathbb{I}_F \mid \mathcal{G}) \mathrm{E}(\mathbb{I}_H \mid \mathcal{G}).$$

Thus, $\mathbb{P}(H \cap F \mid \mathcal{G}) = \mathbb{P}(H \mid \mathcal{G})\,\mathbb{P}(F \mid \mathcal{G})$ almost surely for all $F \in \mathcal{F}$ and $H \in \mathcal{H}$, which allows us to conclude that indeed, $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$.

To prove the reverse implication, we must check that if $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$, then for any $H \in \mathcal{H}$,

$$\int_A \mathbb{P}(H \mid \mathcal{G}) d\mathbb{P} = \int_A \mathbb{P}(H \mid \mathcal{F} \vee \mathcal{G}) d\mathbb{P} = \mathbb{P}(A \cap H) \tag{2.1}$$

holds for all $A \in \mathcal{F} \vee \mathcal{G}$. First, we show that this holds for all $A = F \cap G$, where $G \in \mathcal{G}$ and $F \in \mathcal{F}$. Note that if $A = F \cap G$ with $G \in \mathcal{G}$ and $F \in \mathcal{F}$, then we have

$$\int_A \mathbb{P}(H \mid \mathcal{G}) d\mathbb{P} = \int \mathbb{I}_A \mathrm{E}\left(\mathbb{I}_H \mid \mathcal{G}\right) d\mathbb{P} = \int \mathbb{I}_G \mathbb{I}_F \mathrm{E}\left(\mathbb{I}_H \mid \mathcal{G}\right) d\mathbb{P} = \mathrm{E}\left[\mathrm{E}\left\{\mathbb{I}_G \mathbb{I}_F \mathrm{E}\left(\mathbb{I}_H \mid \mathcal{G}\right) \mid \mathcal{G}\right\}\right].$$

The right-hand term can be rewritten as

$$\mathrm{E}\left\{\mathbb{I}_G \mathrm{E}\left(\mathbb{I}_F \mid \mathcal{G}\right) \mathrm{E}\left(\mathbb{I}_H \mid \mathcal{G}\right)\right\} = \mathrm{E}\left\{\mathbb{I}_G \mathrm{E}\left(\mathbb{I}_{F \cap H} \mid \mathcal{G}\right)\right\} = \int_G \mathbb{P}\left(F \cap H \mid \mathcal{G}\right) d\mathbb{P},$$

where the first inequality holds because $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$. The integral in this last expression satisfies

$$\int_G \mathbb{P}\left(F \cap H \mid \mathcal{G}\right) d\mathbb{P} = \int_G \mathbb{I}_{F \cap H} d\mathbb{P} = \mathbb{P}\left\{G \cap (F \cap H)\right\} = \mathbb{P}\left(A \cap H\right).$$

Note that $\mathcal{C} = \{F \cap G : F \in \mathcal{F} \text{ and } G \in \mathcal{G}\}$ generates $\mathcal{F} \vee \mathcal{G}$. Thus, it suffices to show that $\mathcal{M} = \{A \in \mathcal{F} \vee \mathcal{G} : (2.1) \text{ holds for } A\}$ is a monotone class to show that $\mathcal{M} = \mathcal{F} \vee \mathcal{G}$ since $\mathcal{C} \subset \mathcal{M}$.

Let $A_1 \subset A_2 \subset \cdots$ be an increasing sequence of sets in $\mathcal{M}$ converging to $A = \bigcup_{i=1}^{\infty} A_i$. Then $\mathbb{I}_{A_i} \nearrow \mathbb{I}_A$ as $i \to \infty$, so by monotone convergence, one gets

$$\begin{aligned}\int_A \mathbb{P}\left(H \mid \mathcal{G}\right) d\mathbb{P} &= \lim_{i \to \infty} \int_{A_i} \mathbb{P}\left(H \mid \mathcal{G}\right) d\mathbb{P} \\ &\overset{(2.1)}{=} \lim_{i \to \infty} \int_{A_i} \mathbb{P}\left(H \mid \mathcal{F} \vee \mathcal{G}\right) d\mathbb{P} = \int_A \mathbb{P}\left(H \mid \mathcal{F} \vee \mathcal{G}\right) d\mathbb{P}.\end{aligned}$$

Therefore, $A \in \mathcal{M}$. Alternatively, let $A_1 \supset A_2 \supset \cdots$ be a decreasing sequence of sets in $\mathcal{M}$ converging to $A = \bigcap_{i=1}^{\infty} A_i$. Then, $\mathbb{I}_{A_i} \to \mathbb{I}_A$ as $i \to \infty$ and because $|\mathbb{I}_{A_i}| \leq \mathbb{I}_{A_1}$ for every

integer $i \in \mathbb{N}$, it also follows by dominated convergence that

$$\int_A \mathbb{P}(H \mid \mathcal{G}) d\mathbb{P} = \int_A \mathbb{P}(H \mid \mathcal{F} \vee \mathcal{G}) d\mathbb{P}$$

using the same argument as for increasing sequences. Therefore, $A \in \mathcal{M}$ and so, $\mathcal{M}$ is a monotone class. □

The characterization of conditional independence from Theorem 2.2.1 is much simpler to work with than the original definition. Another important result due to Doob is the chain rule for conditional expectation; see Theorem $8.12$ in Kallenberg (2022). This result pertains to the manipulation of conditional independence statements that involve more than three $\sigma$-fields. Thus, we introduce conditional independence statements that relate to larger numbers of $\sigma$-fields.

For a sequence $\mathcal{F}_1, \mathcal{F}_2, \ldots$ of $\sigma$-fields, let

$$\bigvee_{n=1}^{\infty} \mathcal{F}_n = \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n\right).$$

Then, for any $\sigma$-fields $\mathcal{G}$ and $\mathcal{H}$, the conditional independence statement $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}_1, \mathcal{F}_2, \ldots$ is denoted $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \bigvee_{n=1}^{\infty} \mathcal{F}_n$ and $\mathcal{H} \perp\!\!\!\perp_{\mathcal{F}_1,\ldots,\mathcal{F}_n} \mathcal{F}_{n+1}$ denotes $\mathcal{H} \perp\!\!\!\perp_{\bigvee_{i=1}^{n} \mathcal{F}_i} \mathcal{F}_{n+1}$. With this, we proceed with Doob's chain rule. This proof also follows the arguments from Kallenberg (2022).

**Theorem 2.2.2** (Doob)**.** *For any $\sigma$-fields $\mathcal{G}$ , $\mathcal{H}$ and $\mathcal{F}_1, \mathcal{F}_2, \ldots$, the following are equivalent:*

*(i)* $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}_1, \mathcal{F}_2, \ldots$;

*(ii)* $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G},\mathcal{F}_1,\ldots,\mathcal{F}_n} \mathcal{F}_{n+1}$ *for every* $n \geq 0$.

*Proof.* First, we show that *(i)* implies *(ii)*. If *(i)* holds, then, for every integer $n \in \mathbb{N}$, $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}_1, \ldots, \mathcal{F}_n$ and $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}_1, \ldots, \mathcal{F}_{n+1}$. Given that

$$\mathbb{P}\left(\cdot \mid \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n\right) = \mathbb{P}\left\{\cdot \mid \mathcal{G} \vee \left(\bigvee_{i=1}^{n} \mathcal{F}_i\right)\right\},$$

Theorem 2.2.1 implies that, for all $H \in \mathcal{H}$, one has

$$\mathbb{P}(H \mid \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n) = \mathbb{P}(H \mid \mathcal{G}) = \mathbb{P}(H \mid \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_{n+1})$$

almost surely. Therefore, from the opposite equivalence from Theorem 2.2.1, it follows that $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G},\mathcal{F}_1,\ldots,\mathcal{F}_n} \mathcal{F}_{n+1}$.

Next, we show that *(ii)* implies *(i)*. If *(ii)* holds, then by Theorem 2.2.1, one has that, for every integer $n \geq 0$ and every set $H \in \mathcal{H}$,

$$\mathbb{P}\left(H \mid \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n\right) = \mathbb{P}\left(H \mid \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_{n+1}\right)$$

almost surely. Thus, for any integer $m > 1$, we get

$$\mathbb{P}\left(H \mid \mathcal{G}\right) = \mathbb{P}\left(H \mid \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_m\right) \tag{2.2}$$

So, using the inverse equivalence from Theorem 2.2.1, we get $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}}, \mathcal{F}_1, \ldots, \mathcal{F}_m$ for every integer $m \geq 1$. We will show that this implies $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}_1, \mathcal{F}_2, \ldots$ through a monotone-class argument. Consider

$$\mathcal{A} = \{F : \exists_{N \in \mathbb{N}} \, \forall_{i \in \{1,\ldots,N\}} \, \exists_{F_i \in \mathcal{F}_i} \, F = F_1 \cap \cdots \cap F_n\}$$

and

$$\mathcal{M} = \left\{ F \in \bigvee_{i=1}^{\infty} \mathcal{F}_i : \forall_{H \in \mathcal{H}} \, \mathbb{P}(F \cap H \mid \mathcal{G}) = \mathbb{P}(F \mid \mathcal{G}) \mathbb{P}(H \mid \mathcal{G}) \text{ a.s.} \right\}.$$

Note that $\mathcal{A}$ generates $\bigvee_{i=1}^{\infty} \mathcal{F}_i$. By (2.2), one has

$$\mathbb{P}(H \cap A \mid \mathcal{G}) = \mathbb{P}(H \mid \mathcal{G}) \mathbb{P}(A \mid \mathcal{G})$$

for any $A \in \mathcal{A}$ almost surely, so $\mathcal{A} \subset \mathcal{M}$. Thus, it suffices to show that $\mathcal{M}$ is a $\sigma$-field to show that $\bigvee_{i=1}^{\infty} \mathcal{F}_i = \mathcal{M}$.

Clearly, $\Omega \in \mathcal{M}$. Moreover, let $A \in \mathcal{M}$. Then, almost surely,

$$\begin{aligned}\mathbb{P}(H \cap A^{\complement} \mid \mathcal{G}) &= \mathbb{P}(H \mid \mathcal{G}) - \mathbb{P}(H \cap A \mid \mathcal{G}) = \mathbb{P}(H \mid \mathcal{G}) - \mathbb{P}(H \mid \mathcal{G})\,\mathbb{P}(A \mid \mathcal{G}) \\ &= \mathbb{P}(H \mid \mathcal{G})\,\{1 - \mathbb{P}(A \mid \mathcal{G})\} = \mathbb{P}(H \mid \mathcal{G})\,\mathbb{P}(A^{\complement} \mid \mathcal{G}).\end{aligned}$$

Therefore, $A^{\complement} \in \mathcal{M}$. Finally, let $A_1 \subset A_2 \subset \cdots$ be an increasing sequence in $\mathcal{M}$ converging to $A = \bigcup_{i=1}^{\infty} A_i$. Then, almost surely,

$$\begin{aligned}\mathbb{P}(H \cap A \mid \mathcal{G}) &= \lim_{i \to \infty} \mathbb{P}(H \cap A_i \mid \mathcal{G}) \\ &= \lim_{i \to \infty} \mathbb{P}(H \mid \mathcal{G})\,\mathbb{P}(A_i \mid \mathcal{G}) = \mathbb{P}(H \mid \mathcal{G})\,\mathbb{P}(A \mid \mathcal{G}).\end{aligned}$$

Accordingly, $A \in \mathcal{M}$ and so $\mathcal{M}$ is a $\sigma$-field. □

An important consequence of the chain rule is that $\mathcal{H} \perp\!\!\!\perp_{\mathcal{F}_1,\ldots,\mathcal{F}_n} \mathcal{F}_{n+1}$ for every integer $n \geq 1$ if and only if $\mathcal{H} \perp\!\!\!\perp \mathcal{F}_1, \mathcal{F}_2, \ldots$ This proof also implies that for every integer $n \geq 2$, $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}_1 \ldots, \mathcal{F}_n$ if and only if $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G},\mathcal{F}_1,\ldots,\mathcal{F}_m} \mathcal{F}_{m+1}$ for all $m \in \{0, \ldots, n-1\}$.

## 2.3 Conditional independence assumptions

Aggregation-tree models always provide a unique distribution for all the node variables in the tree structure. However, as previously mentioned, it is possible to obtain a unique and well-defined joint distribution for the leaves of the tree structure if certain assumptions about the conditional independence of the nodes are satisfied. We will present a more general version of these assumptions that apply to tree structures where more than two variables can be aggregated at each step, as some of these proofs are significantly different when nodes are allowed to have multiple sibling nodes in the tree structure. Originally, Arbenz et al. (2012) used the following conditional independence assumption.

**Conditional Independence Assumption (A)**: For any branching node $M_I$, $I \in \mathcal{B}(\mathcal{T})$, with descendants $\{M_J : J \in \mathcal{D}(I)\}$, one has $\{M_J : J \in \mathcal{D}(I)\} \perp\!\!\!\perp_{M_I} \{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\}$.

Côté & Genest (2015) provided a similar conditional independence assumption for the models they presented. Their condition is stated next.

**Conditional Independence Assumption (B)**: For any branching node $M_I$, $I \in \mathcal{B}(\mathcal{T})$, with leaf descendants $\{M_J : J \in \mathcal{L}(I)\} = \{X_i : i \in I\}$, one has $\{X_i : i \in I\} \perp\!\!\!\perp_{M_I} \{X_i : i \in \bar{I}\}$.

However, we will show in the sequel that a weaker form of conditional independence assumption suffices.

**Local Conditional Independence Assumption (C)**: For any branching node $M_I$, $I \in \mathcal{B}(\mathcal{T})$, with children nodes $M_{I_1}, \ldots, M_{I_n}$, we assume that for each $k \in \{1, \ldots, n\}$, the leaf descendants of each $M_{I_k}$ are conditionally independent of the leaf descendants of all its sibling nodes conditioning on $M_{I_k}$, i.e., $\{X_j : j \in I_k\} \perp\!\!\!\perp_{M_{I_k}} \{X_j : j \in \bigcup_{i \neq k} I_i\}$.

We will prove below that all of these conditional independence assumptions are equivalent. If $(M_I)_{I \in \mathcal{T}}$ satisfies any of the assumptions (A), (B) or (C), then it must satisfy the other two. This result is formally stated and proved below. First, we proceed by proving the equivalence of (A) and (B).

**Theorem 2.3.1.** *Given an aggregation-tree model for a random vector $(X_1, \ldots, X_d)$ with a tree structure $\mathcal{T}$ and aggregation function $a$, then $(M_I)_{I \in \mathcal{T}}$ satisfies the conditional independence assumption (A) if and only if it satisfies the conditional independence assumption (B).*

*Proof.* It is clear that if $(M_I)_{I \in \mathcal{T}}$ satisfies the conditional independence assumption (A), then it satisfies the conditional independence assumption (B) because $\{X_i : i \in I\} \subset \{M_J : J \in \mathcal{D}(I)\}$ and $\{X_i : i \in \bar{I}\} \subset \{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\}$. Hence, it remains to show that (B) implies (A).

Let $I \in \mathcal{B}(\mathcal{T})$. For any $J \in \mathcal{D}(I)$, one has $\sigma(M_J) \subset \sigma(X_i : i \in I)$. Thus, $\sigma(M_J : J \in \mathcal{D}(I))$ is a sub-field of $\sigma(X_i : i \in I)$. As for $\sigma\{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\}$, we may rewrite this $\sigma$-field as $\mathcal{X}_I \vee \mathcal{Y}_I$, where

$$\mathcal{X}_I = \sigma(M_J : J \in \mathcal{T} \text{ and } I \subsetneq J), \quad \mathcal{Y}_I = \sigma(M_J : J \in \mathcal{T} \text{ and } I \cap J = \emptyset).$$

Given that if $I \cap J = \emptyset$, then $J \subset \bar{I}$, one has $\mathcal{Y}_I \subset \sigma(X_i : i \in \bar{I})$. When $I \subsetneq J$, $M_J$ is a measurable function of $M_I$ and $\{X_i : i \in J \setminus I\} \subset \{X_i : i \in \bar{I}\}$. Thus, $\mathcal{X}_I$ is a sub-field of $\sigma(M_I) \vee \sigma(X_i : i \in \bar{I})$. Rewriting the conditional independence statement $\{M_J : J \in \mathcal{D}(I)\} \perp\!\!\!\perp_{M_I} \{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\}$ in terms of $\sigma$-fields, we get

$$\sigma\{M_J : J \in \mathcal{D}(I)\} \perp\!\!\!\perp_{\sigma(M_I)} \sigma\{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\}.$$

Given that $\sigma\{M_J : J \in \mathcal{D}(I)\} \subset \sigma\{X_i : i \in I\}$ and $\sigma\{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\} \subset \{\sigma(X_i : i \in \bar{I}) \vee \sigma(M_I)\}$, we know that the previous statement is a consequence of the following conditional independence statement:

$$\sigma(X_i : i \in I) \perp\!\!\!\perp_{\sigma(M_I)} \{\sigma(X_i : i \in \bar{I}) \vee \sigma(M_I)\}.$$

By Doob's chain rule (Theorem 2.2.2), this statement is equivalent to

$$\sigma(X_i : i \in I) \perp\!\!\!\perp_{\sigma(M_I)} \sigma(X_i : i \in \bar{I}),$$

which is exactly the $\sigma$-field form of $\{X_i : i \in I\} \perp\!\!\!\perp_{M_I} \{X_i : i \in \bar{I}\}$ from (B). Therefore, $\{X_i : i \in I\} \perp\!\!\!\perp_{M_I} \{X_i : i \in \bar{I}\}$ implies $\{M_J : J \in \mathcal{D}(I)\} \perp\!\!\!\perp_{M_I} \{M_J : J \in \mathcal{T} \setminus \mathcal{D}(I)\}$. As a consequence, if $(M_I)_{I \in \mathcal{T}}$ satisfies (B) then it satisfies (A). □

Now that the equivalence of (A) and (B) has been established, we prove the equivalence of (B) and (C) to show that all the conditional independence assumptions are equivalent.

**Theorem 2.3.2.** *Given an aggregation-tree model for random vector $(X_1, \ldots, X_d)$ with a tree structure $\mathcal{T}$ and aggregation function $a$, then $(M_I)_{I \in \mathcal{T}}$ satisfies the local conditional independence assumption (C) if and only if it satisfies the conditional independence assumption (B).*

Before we prove this result, we require an important lemma.

**Lemma 2.3.3.** *For any $\sigma$-fields $\mathcal{A}$, $\mathcal{G}$ ,$\mathcal{F}$, $\mathcal{H}$ such that $\mathcal{G} \subset \mathcal{F} \subset \mathcal{H}$, if $\mathcal{A} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$, then $\mathcal{A} \perp\!\!\!\perp_{\mathcal{F}} \mathcal{H}$.*

*Proof.* To prove this result, it is sufficient to show that $\mathbb{P}(A \mid \mathcal{G}) = \mathbb{P}(A \mid \mathcal{F})$ almost surely for any $A \in \mathcal{A}$. This is because if $\mathcal{A} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$, then $\mathbb{P}(A \mid \mathcal{H} \vee \mathcal{G}) = \mathbb{P}(A \mid \mathcal{G})$ almost surely by Theorem 2.2.1. However, given that $\mathcal{G} \subset \mathcal{F} \subset \mathcal{H}$, one has

$$\mathbb{P}(A \mid \mathcal{H} \vee \mathcal{G}) = \mathbb{P}(A \mid \mathcal{H}) \quad \text{and} \quad \mathbb{P}(A \mid \mathcal{H} \vee \mathcal{F}) = \mathbb{P}(A \mid \mathcal{H})$$

almost surely. Thus, if $\mathbb{P}(A \mid \mathcal{G}) = \mathbb{P}(A \mid \mathcal{F})$ almost surely, we have

$$\mathbb{P}(A \mid \mathcal{H} \vee \mathcal{F}) = \mathbb{P}(A \mid \mathcal{H}) = \mathbb{P}(A \mid \mathcal{H} \vee \mathcal{G}) = \mathbb{P}(A \mid \mathcal{G}) = \mathbb{P}(A \mid \mathcal{F})$$

almost surely, which implies $\mathcal{A} \perp\!\!\!\perp_{\mathcal{F}} \mathcal{H}$ by Theorem 2.2.1.

Let $F \in \mathcal{F}$, then $\mathbb{P}(A \cap F) = \mathrm{E}\{\mathbb{P}(A \mid \mathcal{F})\mathbb{I}_F\}$. By the tower property,

$$\mathrm{E}\{\mathbb{P}(A \mid \mathcal{F})\mathbb{I}_F\} = \mathrm{E}\big[\mathrm{E}\{\mathrm{E}\left(\mathbb{I}_{\mathrm{A}} \mid \mathcal{H}\right) \mid \mathcal{F}\}\mathbb{I}_{\mathrm{F}}\big].$$

Given that $\mathrm{E}(\mathbb{I}_A \mid \mathcal{H}) = \mathrm{E}(\mathbb{I}_A \mid \mathcal{G})$ almost surely and $\mathrm{E}(\mathbb{I}_A \mid \mathcal{G})$ is $\mathcal{F}$-measurable because $\mathcal{G} \subset \mathcal{F}$, we get

$$\mathrm{E}\{\mathbb{P}(A \mid \mathcal{F})\mathbb{I}_F\} = \mathrm{E}[\mathrm{E}\{\mathrm{E}(\mathbb{I}_A \mid \mathcal{G}) \mid \mathcal{F}\}\mathbb{I}_F] = \mathrm{E}\{\mathrm{E}(\mathbb{I}_A \mid \mathcal{G})\mathbb{I}_F\} = \mathrm{E}\{\mathbb{P}(A \mid \mathcal{G})\mathbb{I}_F\}.$$

Thus, $\mathbb{P}(A \cap F) = \mathrm{E}\{\mathbb{P}(A \mid \mathcal{G})\mathbb{I}_F\}$ for all $F \in \mathcal{F}$. By the almost-sure uniqueness of conditional probability, this implies that $\mathbb{P}(A \mid \mathcal{G}) = \mathbb{P}(A \mid \mathcal{F})$ almost surely. □

Intuitively, this result asserts that incorporating more information in the conditioning does not affect the conditional independence. We now proceed with the proof of Theorem 2.3.1.

*Proof.* Clearly, conditional independence assumption (B) implies conditional independence assumption (C), so we will show that (C) implies (B). Suppose that $(M_I)_{I \in \mathcal{T}}$ sat-

isfies (C). We will prove inductively that for every branching node $M_I$, $I \in \mathcal{B}(I)$, we have $\{X_\ell : \ell \in I\} \perp\!\!\!\perp_{M_I} \{X_\ell : \ell \in \bar{I}\}$.

**Base case** ($X_I$ is a child of the root node): Let $|\mathcal{B}(\mathcal{T})| > 1$ and let $M_{I_1}, \ldots, M_{I_n}$ be the children nodes of the root of the tree $M_D = a(X_i : i \in \{1, \ldots, d\})$. By (C), for any $M_{I_j}$, we have $\{X_\ell : \ell \in I_j\} \perp\!\!\!\perp_{M_{I_j}} \{X_\ell : \ell \in \bigcup_{k \neq j} I_k\}$. However, given that $M_{I_j}$ is a child of the root node, $\{X_\ell : \ell \in \bigcup_{k \neq j} I_k\} = \{X_\ell : \ell \in \bar{I}_j\}$. Therefore, $\{X_\ell : \ell \in I_j\} \perp\!\!\!\perp_{M_{I_j}} \{X_\ell : \ell \in \bar{I}_j\}$.

**Induction step**: Let $M_{I_1}, \ldots, M_{I_n}$ be the children nodes of a branching node $M_I \neq M_D$ and assume that $\{X_\ell : \ell \in I\} \perp\!\!\!\perp_{M_I} \{X_\ell : \ell \in \bar{I}\}$. Rewriting this conditional independence statement in terms of $\sigma$-fields, we get $\sigma(X_i : i \in I) \perp\!\!\!\perp_{\sigma(M_I)} \sigma(X_i : i \in \bar{I})$, where $\sigma(M_J : J \in A) = \bigvee_{J \in A} \sigma(M_J)$. Given that $M_I$ is a measurable function of its children nodes, we have

$$\sigma\left(M_I\right) \subset \bigvee_{k=1}^{n} \sigma(M_{I_k}) \subset \sigma(M_{I_j}) \bigvee \left\{ \bigvee_{k \neq j} \sigma(X_i : i \in I_k) \right\} \subset \sigma(X_i : i \in I).$$

Fix an arbitrary $j \in \{1, \ldots, n\}$. By Lemma 2.3.3, $\sigma(X_i : i \in I) \perp\!\!\!\perp_{\sigma(M_I)} \sigma(X_i : i \in \bar{I})$ implies

$$\sigma(X_i : i \in I) \perp\!\!\!\perp_{\sigma(M_{I_j}) \vee (\bigvee_{k \neq j} \sigma(X_i : i \in I_k))} \sigma(X_i : i \in \bar{I}).$$

In particular,

$$\sigma(X_i : i \in I_j) \perp\!\!\!\perp_{\sigma(M_{I_j}) \vee (\bigvee_{k \neq j} \sigma(X_i : i \in I_k))} \sigma(X_i : i \in \bar{I}).$$

Moreover, $\sigma(X_i : i \in I_j) \perp\!\!\!\perp_{\sigma(M_{I_j})} \bigvee_{k \neq j} \sigma(X_i : i \in I_k)$ by (C). Thus, by Doob's chain rule Theorem 2.2.2, we deduce that

$$\sigma(X_i : i \in I_j) \perp\!\!\!\perp_{\sigma(M_{I_j})} \sigma(X_i : i \in \bar{I}) \bigvee \left\{ \bigvee_{k \neq j} \sigma(X_i : i \in I_k) \right\}.$$

Rewriting this conditional independence statement in terms of random variables gives $\{X_\ell : \ell \in I_j\} \perp\!\!\!\perp_{M_{I_j}} \{X_\ell : \ell \in \bar{I}_j\}$, as desired. This argument can be repeated for every $I \in \mathcal{B}(\mathcal{T})$ by going lower down the hierarchy of the tree structure. $\square$

Finally, we may show that under any of these conditional independence assumptions, the model construction provides a unique and well-defined joint distribution for the leaf nodes of the tree. We choose to work with (C) for this proof.

**Theorem 2.3.4.** *Given an aggregation-tree model for random vector $(X_1, \ldots, X_d)$ with a tree structure $\mathcal{T}$, aggregation function $a$, and marginal distributions $F_1, \ldots, F_d$. If $(M_I)_{I \in \mathcal{T}}$ satisfies the conditional independence assumption (C), a unique and well-defined joint distribution for $(X_1, \ldots, X_d)$ exists.*

*Proof.* By construction $X_1, \ldots, X_d$ are the leaf nodes of the tree structure. It suffices to prove that a joint cumulative distribution function on $\mathbb{R}^d$ with $\boldsymbol{x} \mapsto \mathbb{P}(X_j \leq x_j : j \in \{1, \ldots, d\})$ exists and is unique. To do this, we will inductively prove that for every branching node $M_I$, $\mathbb{P}(X_j \leq x_j : j \in I)$ is well -defined and unique.

**Base case** ($I \in \mathcal{L}(\mathcal{T})$): If $I \in \mathcal{L}(\mathcal{T})$, then $M_I = X_i$ for some $i \in \{1, \ldots, d\}$, therefore $\mathcal{L}(I) = \{\{i\}\}$. By the model assumptions, we have that, for all $x_1, \ldots, x_d \in \mathbb{R}$,

$$\mathbb{P}(X_j \leq x_j : j \in I) = \mathbb{P}(X_i \leq x_i) = F_i(x_i).$$

**Induction step**: Let $I \in \mathcal{B}(\mathcal{T})$ and let $M_{I_1}, \ldots, M_{I_n}$ be the children nodes of $M_I$. We assume that $\mathbb{P}\{X_j \leq x_j : j \in I_k\}$ is unique and well-defined for each $x_j \in \mathbb{R}$, $j \in I_k$ and each of the children nodes of $M_I$. To this end, fix an arbitrary $x_i \in \mathbb{R}$, $i \in \bigcup_{k=1}^{n} I_k$. We rewrite the joint distribution using iterated conditioning on $M_{I_1}, \ldots, M_{I_n}$, viz.

$$\mathbb{P}\left(X_j \leq x_j : j \in I\right) = \mathrm{E}\big\{\mathbb{P}(X_j \leq x_j : j \in I \mid M_{I_1}, \ldots, M_{I_n})\big\}.$$

We claim that

$$\mathbb{P}(X_j \leq x_j : j \in I \mid M_{I_1}, \ldots, M_{I_n}) = \prod_{k=1}^{n} \mathbb{P}(X_j \leq x_j : j \in I_k \mid M_{I_1}, \ldots, M_{I_n}).$$

To establish this claim, we prove that $\mathcal{B}_k \perp\!\!\!\perp_{\mathcal{A}} \mathcal{B}_{(k)}$, where for each $k \in \{1, \ldots, n\}$,

$$\mathcal{B}_k = \bigvee_{j \in I_k} \sigma(X_j), \quad \mathcal{B}_{(k)} = \bigvee_{j \neq k} \mathcal{B}_j, \quad \mathcal{A}_k = \sigma(M_{I_k}), \quad \mathcal{A} = \bigvee_{j=1}^{n} \mathcal{A}_j, \quad \mathcal{A}_{(k)} = \bigvee_{j \neq k} \mathcal{A}_j.$$

For any $k \in \{1, \ldots, n\}$, let $B_k \in \mathcal{B}_k$. We know that $\mathcal{B}_k \perp\!\!\!\perp_{\mathcal{A}_k} \mathcal{B}_{(k)}$ by (C). Also, for each $j \neq k$, $\mathcal{A}_j \subset \mathcal{B}_{(k)}$ because each $M_{I_j}$ is a measurable function of its leaf descendants. Therefore, $\mathcal{B}_k \perp\!\!\!\perp_{\mathcal{A}_k} \mathcal{B}_{(k)} \bigvee \mathcal{A}_{(k)}$. Thus, we have

$$\mathbb{P}\left(B_k \mid \mathcal{B}_{(k)} \vee \mathcal{A}\right) = \mathbb{P}\left(B_k \mid \mathcal{B}_{(k)} \vee \mathcal{A}_k \vee \mathcal{A}_{(k)}\right) = \mathbb{P}\left(B_k \mid \mathcal{A}_k\right)$$

almost surely by the $(i) \implies (ii)$ implication from Theorem 2.2.1. Moreover, given that $\mathcal{A}_{(k)}$ is a sub-field of $\mathcal{B}_{(k)} \bigvee \mathcal{A}_{(k)}$, $\mathcal{B}_k \perp\!\!\!\perp_{\mathcal{A}_k} \mathcal{B}_{(k)} \bigvee \mathcal{A}_{(k)}$ also implies $\mathcal{B}_k \perp\!\!\!\perp_{\mathcal{A}_k} \mathcal{A}_{(k)}$.

By the $(i) \implies (ii)$ implication from Theorem 2.2.1 once again, we get $\mathbb{P}\left(B_k \mid \mathcal{A}_k\right) = \mathbb{P}\left(B_k \mid \mathcal{A}_k \vee \mathcal{A}_{(k)}\right) = \mathbb{P}\left(B_k \mid \mathcal{A}\right)$. This implies that $\mathbb{P}\left(B_k \mid \mathcal{B}_{(k)} \vee \mathcal{A}\right) = \mathbb{P}\left(B_k \mid \mathcal{A}\right)$ for all $B_k \in \mathcal{B}_k$. Consequently, by the $(ii) \implies (i)$ implication from Theorem 2.2.1, $\mathcal{B}_k \perp\!\!\!\perp_{\mathcal{A}} \mathcal{B}_{(k)}$ and given that this holds for all $k \in \{1, \ldots, n\}$, one finds

$$\mathbb{P}\left(X_j \leq x_j : j \in I \mid M_{I_1}, \ldots, M_{I_n}\right) = \prod_{k=1}^{n} \mathbb{P}\left(X_j \leq x_j : j \in I_k \mid M_{I_1}, \ldots, M_{I_n}\right).$$

Moreover, for every $k \in \{1, \ldots, n\}$,

$$\mathbb{P}(X_j \leq x_j : j \in I_k \mid M_{I_1}, \ldots, M_{I_n}) = \mathbb{P}(X_j \leq x_j : j \in I_k \mid M_{I_k}),$$

once again by Theorem 2.2.1, because $(X_j : j \in I_k) \perp\!\!\!\perp_{M_{I_k}} (M_{I_j} : j \neq k)$ by conditional independence assumption (C). Combining this with the previous result, we have

$$\mathbb{P}\left(X_j \leq x_j : j \in I\right) = \mathrm{E}\left\{\prod_{k=1}^{n} \mathbb{P}\left(X_j \leq x_j : j \in I_k \mid M_{I_k}\right)\right\}.$$

By the induction hypothesis, then each $\mathbb{P}(X_j \leq x_j : j \in I_k \mid M_{I_k})$ is unique and well-defined. Under the model construction, the joint distribution of the vector $(M_{I_1}, \ldots, M_{I_n})$ is given by the copula $C_{I_1,\ldots,I_n} = C$ and the marginal distributions $F_{I_1}, \ldots, F_{I_n}$. By Sklar's theorem, the random vector $(M_{I_1}, \ldots, M_{I_n})$ has the same distribution as $(F_{I_1}^{-1}(U_1), \ldots, F_{I_n}^{-1}(U_n))$, where the vector $(U_1, \ldots, U_n)$ has distribution $C$. Thus,

$$\mathbb{P}\left(X_j \leq x_j : j \in I\right) = \int_{[0,1]^n} \prod_{k=1}^{n} \mathbb{P}\left\{X_j \leq x_j : j \in I_k \mid M_{I_k} = F_{I_k}^{-1}(u_k)\right\} dC(u).$$

This gives a unique and well-defined formula for $\mathbb{P}(X_j \leq x_j : j \in I)$. Similarly to the previous proof, this argument can be repeated for every branching node $M_I$, $I \in \mathcal{B}(\mathcal{T})$, to show that the joint distribution of $(X_j : j \in I)$ is uniquely defined by going up on the hierarchy of the tree structure. □

However, we shall soon illustrate that, in general, there is no closed form for the joint pdf of aggregation-tree models. Thus, to gain further insight into the properties of these models, we examine sampling techniques for aggregation-tree models.

# Chapter 3

# Sampling from aggregation-tree models

In this chapter, we present two approaches to sample from an aggregation-tree model which uses aggregation function $a$, copulas $C_1, \ldots, C_{d-1}$, and marginal distributions $F_1, \ldots, F_d$. One of these approaches, the direct simulation approach, is based on simulating directly from the joint distribution, provided that the conditional independence assumption (C) holds, using conditional copulas. The second approach, the sample reordering approach, consists of reordering independent samples drawn from the marginal distributions $F_1, \ldots, F_d$ to introduce the dependence structures specified by the aggregation-tree model with copulas $C_1, \ldots, C_{d-1}$. These sampling techniques are required for the application of certain model validation techniques which we will introduce later on.

## 3.1 Direct simulation approach

Under the conditional independence assumption (C), the model will provide a unique joint distribution for the random vector $(X_1, \ldots, X_d)$. If we are able to sample from this joint distribution, we can compute any remaining branching nodes in the tree. We can obtain a recursive formula for the joint density of all the leaf nodes of the tree. The following proof follows the same approach as that which was presented in Côté & Genest

(2015), where partial sums were used as the aggregation function. However, ours is valid for an arbitrary aggregation function.

**Theorem 3.1.1.** *Suppose that an aggregation-tree model is given for the variables $X_1, \ldots, X_d$ with tree structure $\mathcal{T}$, where for each $i \in \{1, \ldots, d-1\}$, the variables $M_{B_{1i}} = M_{A_{\ell_i i}}$ and $M_{B_{2i}} = M_{A_{k_i i}}$ are joined at the $i$th aggregation step. If $(M_I)_{I \in \mathcal{T}}$ satisfies the conditional independence assumption, then the joint density for $\boldsymbol{X}_D = (X_1, \ldots, X_d)$ is given by*

$$\mathbf{f}_D(x_1, \ldots, x_d) = \prod_{i=1}^{d-1} c_{B_{1i}B_{2i}}[F_{B_{1i}}\{a(x_j : j \in B_{1i})\}, F_{B_{2i}}\{a(x_j : j \in B_{2i})\}] \prod_{i=1}^{d} f_i(x_i), \qquad (3.1)$$

*where each $c_{B_{1i}B_{2i}}$ is the density of the copula $C_{B_{1i}B_{2i}}$ that joins $M_{B_{1i}}$ and $M_{B_{2i}}$, and each $f_i$ is the density function of $F_i$.*

*Proof.* We will proceed by induction on $d \in \mathbb{N}$.

**Base case** ($d = 2$): The only possible tree structure for $d = 2$ variables has $A_{11} = \{1\}$, $A_{21} = \{2\}$. Assume that $\mathbf{F}_{\{1,2\}} = C_1(F_1, F_2)$. It follows that one has, for all $x_1, x_2 \in \mathbb{R}$,

$$\mathbf{f}_{\{1,2\}}(x_1, x_2) = c_1\{F_1(x_1), F_2(x_2)\} f_1(x_1) f_2(x_2),$$

which is of the form (3.1).

**Induction step** ($d > 2$): Suppose that the statement of the theorem is true for any aggregation-tree model of dimension $k < d+1$. We will show that Eq. (3.1) also holds for any tree structure for $d+1$ variables. Consider a tree structure $\mathcal{T}$ for $d+1$ variables where we have $\mathcal{T}_d = \{A_{1d}, A_{2d}\}$. First, we consider the case where $|A_{1d}| > 1$ and $|A_{2d}| > 1$. The sub-tree obtained from the nodes descendants of $M_{A_{1d}}$ with it as the root node represents a tree structure with $|A_{1d}| = d_1 \leq d$ leaf nodes. By the conditional independence assumption (C), if $\boldsymbol{x}_A = (x_\ell : \ell \in A)$ and $m_A = a(\boldsymbol{x}_A)$, then

$$\mathbf{f}_D(x_1, \ldots, x_{d+1}) = \mathbf{f}_{A_{1d}}(\boldsymbol{x}_{A_{1d}} \mid M_{A_{1d}} = m_{A_{1d}}) \mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}) f_{M_{A_{1d}}}(m_{A_{1d}}),$$

where $f_{M_{A_{1d}}}$ is the density for $M_{A_{1d}}$. Given that $d_1 \leq d$, one can invoke the induction hypothesis to get

$$\begin{aligned}\mathbf{f}_{A_{1d}}(\boldsymbol{x}_{A_{1d}} \mid M_{A_{1d}} = m_{A_{1d}}) f_{M_{A_{1d}}}(m_{A_{1d}}) &= \mathbf{f}_{A_{1d}}(\boldsymbol{x}_{A_{1d}}) \\ &= \prod_{i \in \tilde{D}_{1d}} c_{B_{1i}B_{2i}}\{F_{B_{1i}}(m_{B_{1i}}), F_{B_{2i}}(m_{B_{2i}})\} \prod_{j \in A_{1d}} f_j(x_j), \quad (3.2)\end{aligned}$$

where $\tilde{D}_{1d}$ is the set of indices for the copulas among $C_{B_{11}B_{21}}, \ldots, C_{B_{1(d-1)}B_{2(d-1)}}$ that were used to join the descendants of $M_{A_{1d}}$ in the tree structure. Moreover, given that the model specifies that $C_{A_{1d}A_{2d}}$ is the copula of $(M_{A_{1d}}, M_{A_{2d}})$ with density $c_{A_{1d}A_{2d}}$, we have

$$f_{M_{A_{2d}}}(m_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}) = c_{A_{1d}A_{2d}}\{F_{A_{1d}}(m_{A_{1d}}), F_{A_{2d}}(m_{A_{2d}})\} f_{M_{A_{2d}}}(m_{A_{2d}}). \quad (3.3)$$

We may also apply the induction hypothesis for the joint distribution of $(X_i : i \in A_{2d})$. By the conditional independence assumption (C), we deduce that

$$\begin{aligned}&\mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}) \\ &\quad = \mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}, M_{A_{2d}} = m_{A_{2d}}) f_{M_{A_{2d}}}(m_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}) \\ &\quad = \mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}} \mid M_{A_{2d}} = m_{A_{2d}}) f_{M_{A_{2d}}}(m_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}).\end{aligned}$$

From (3.3), we get

$$\mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}) = \mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}}) c_{A_{1d}A_{2d}}\{F_{A_{1d}}(m_{A_{1d}}), F_{A_{2d}}(m_{A_{2d}})\}.$$

Using the induction hypothesis, we get

$$\mathbf{f}_{A_{2d}}(\boldsymbol{x}_{A_{2d}} \mid M_{A_{1d}} = m_{A_{1d}}) = \prod_{i \in \tilde{D}_{2d}} c_{B_{1i}B_{2i}}\{F_{B_{1i}}(m_{B_{1i}}), F_{B_{2i}}(m_{B_{2i}})\} \prod_{j \in A_{2d}} f_j(x_j).$$

Combining this result with (3.2), we obtain the desired formula.

If $A_{2d} = \{x_j\}$ for some $j \in \{1, \ldots, d+1\}$, the situation is simpler. Indeed, the model assumes that

$$f_j(x_j \mid M_{A_{1d}} = m_{A_{1d}}) = c_{A_{1d}A_{2d}}\{F_{A_{1d}}(m_{A_{1d}}), F_j(x_j)\} f_j(x_j).$$

Combining this with (3.2) yields the desired result. □

This result provides a formula for the joint density for $(X_1, \ldots, X_d)$ when bivariate copulas are exclusively used. The modeler may then select their preferred method to obtain a sample from the density. The main drawback for this approach is that the formula for the density may not have a closed form. This may depend on the choice of the aggregation function for the model.

For example, even when working with partial sums for the aggregation function, the distribution functions of the branching nodes in the tree are given by convolutions which may not have a closed form. Moreover, the formula for the joint density is only valid when the conditional densities $\mathbf{f}_{A_{1i}}(\boldsymbol{x}_{A_{1i}}|M_{A_{1i}})$ exist for every integer $i \in \{1, \ldots, d-1\}$. Therefore, we may be unable to use these methods for models that use certain choices of aggregation function such as $a(x, y) = \max(x, y)$, where the conditional distribution of $(X_j : j \in A_{1i})$ given $M_{A_{1i}} = \max(X_j : j \in A_{1i})$ is not absolutely continuous.

As we are interested in using maxima for aggregation-tree models in the extreme-value setting, we need an alternative sampling method. We suggest the alternative approach which uses sample reordering because this approach works regardless of the choice of aggregation function, copulas or marginal distributions.

The reordering method presented by Iman & Conover (1992) can be generalized to create dependent data with a more varied range of dependence structures. To sample from a continuous multivariate distribution with marginal distributions $F_1, \ldots, F_d$ and a copula $C$, consider a sample $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$, where for each $i \in \{1, \ldots, n\}$, the random vector $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{id})$ is drawn from the copula $C$.

Consider a matrix $\boldsymbol{X} \in \mathbb{R}^{d \times n}$, where the rows $\boldsymbol{X}_{1.}, \ldots, \boldsymbol{X}_{d.}$ of $\boldsymbol{X}$ are mutually independent samples $\boldsymbol{X}_{i.} = (X_{i1}, \ldots, X_{in})$ from the distribution function $F_i$. Let $R_{ij}$ be the rank of $U_{ij}$ among $U_{1j}, \ldots, U_{nj}$, and consider the permutations $\pi_j$ where $\pi_i(j) = r$ if $R_{ij} = r$. We reorder the rows of $\boldsymbol{X}$ and obtain $\tilde{\boldsymbol{X}}$ such that $\tilde{\boldsymbol{X}}_{ij} = X_{i(\pi_i(j))}$, where $X_{i(1)} < \cdots < X_{i(n)}$.

With this permutation of the rows of the matrix $\boldsymbol{X}$ containing the independent samples, the ranks of each reordered sample $\tilde{\boldsymbol{X}}_{i1}, \ldots, \tilde{\boldsymbol{X}}_{in}$ matches exactly the ranks of $U_{1i}, \ldots$, $U_{ni}$ for each $i \in \{1, \ldots, d\}$. Consequently, given $\tilde{\boldsymbol{X}}_i = (\tilde{X}_{1i}, \ldots, \tilde{X}_{di})$, the empirical copula for $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n$ will be the same as the empirical copula for $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$.

Thus, $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n$ constitutes a random sample from a joint distribution with empirical marginal distributions $\hat{F}_i$ obtained from $\tilde{X}_{i1}, \ldots, \tilde{X}_{in}$. The latter is close to $F_i$ because each set $\{X_{i1}, \ldots, X_{in}\}$ was sampled from $F_i$; moreover, the empirical copula is close to $C$ because $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ were sampled from $C$.

The original method presented in Iman & Conover (1992) is a special case of this approach where the independent samples were drawn from Gaussian distributions and reordered to match the ranks of a sample drawn from a Gaussian copula with correlation matrix $R$. Using this reordering approach, one can sample from any joint distribution given that one can generate independent samples from the specified marginal distributions and copulas. These methods can be further adjusted to sample from an aggregation-tree model.

## 3.2 The recursive Iman–Conover algorithm

Both Arbenz et al. (2012) and Côté & Genest (2015) suggest a recursive reordering method where given a tree structure $\mathcal{T}$, copulas $C_1, \ldots, C_{d-1}$, and marginal distributions $F_1, \ldots, F_d$, we obtain a multivariate sample for $\boldsymbol{X} = (X_1, \ldots, X_d)$, where the sample for each of the $X_i$ comes from the marginal distribution $F_i$ and for all $j \in \{1, \ldots, d-1\}$, and if the dependence structure of $(X_{B_{1j}}, X_{B_{2j}})$ is modeled by the copula $C_j$, the empirical copula from the sample for $(X_{B_{1j}}, X_{B_{2j}})$ is close to $C_j$.

**Algorithm**: We may use the following algorithm to generate a sample of size $n$ from a specified aggregation-tree model with aggregation function $a$.

*Step 1*: Let $N > n$. For each $i \in \{1, \ldots, d\}$, generate a sample $x_i^{(1)}, \ldots, x_i^{(N)}$ from the distribution $F_i$. We store the samples along the rows of a matrix. Let $\boldsymbol{X} \in \mathbb{R}^{d \times N}$ with $\boldsymbol{X}_{ij} = x_i^{(j)}$ for all $i \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, N\}$.

*Step 2*: For each $j \in \{1, \ldots, d\}$, generate a sample $(U_j^{(1)}, V_j^{(1)}), \ldots, (U_j^{(N)}, V_j^{(N)})$ of size $N$ from the copula $C_j$ such that $U_j^{(1)} < \cdots < U_j^{(N)}$. Moreover, define permutations $p_1, \ldots, p_{d-1}$ of $1, \ldots, N$ where $p_j(s) = r$ if $V_j^{(s)}$ has rank $r$ among $V_j^{(1)}, \ldots, V_j^{(N)}$.

*Step 3*: For each $j \in \{1, \ldots, d-1\}$, we compute $m_{B_{1j}}^{(s)} = a(x_i^{(s)} : i \in B_{1j})$ and $m_{B_{2j}}^{(s)} = a(x_i^{(s)} : i \in B_{2j})$. Define the permutations $\tilde{p}_{1,1}, \ldots, \tilde{p}_{d-1,1}, \tilde{p}_{1,2}, \ldots, \tilde{p}_{d-1,2}$ of $\{1, \ldots, N\}$, where $\tilde{p}_{j,1}(s) = r$ if $m_{B_{1j}}^{(s)}$ has rank $r$ among $m_{B_{1j}}^{(1)}, \ldots, m_{B_{1j}}^{(N)}$ and $\tilde{p}_{j,2}(s) = r$ if $m_{B_{2j}}^{(s)}$ has rank $r$ among $m_{B_{2j}}^{(1)}, \ldots, m_{B_{2j}}^{(N)}$.

*Step 4*: We reorder the rows of $\boldsymbol{X}$ as follows. First, we reorder the rows that are the components of $m_{B_{1j}}$ and $m_{B_{2j}}$ so that they match the ranks of $m_{B_{1j}}^{(1)}, \ldots, m_{B_{1j}}^{(N)}$ and $m_{B_{2j}}^{(1)}, \ldots, m_{B_{2j}}^{(N)}$. Thus,

$$\tilde{p}_j(\boldsymbol{X})_{ik} = \begin{cases} x_{i\tilde{p}_{j,1}(k)} & \text{if } k \in B_{1j}, \\ x_{i\tilde{p}_{j,2}(k)} & \text{if } k \in B_{2j}, \\ x_{ik} & \text{otherwise.} \end{cases}$$

This is so that the order of the components of $m_{B_{1j}}$ and $m_{B_{2j}}$ coincide with their ranks. Afterwards, the rows that correspond to the components of $m_{B_{1j}}$ and $m_{B_{2j}}$ are reordered so that they have the same ranks as the sample from the copula $C_j$.

That is,

$$p_j(\boldsymbol{X})_{ik} = \begin{cases} x_{ip_j(k)} & \text{if } k \in B_{2j}, \\ x_{ik} & \text{otherwise.} \end{cases}$$

We recursively define $\boldsymbol{X}^{(j)}$ by setting $\boldsymbol{X}^{(0)} = \boldsymbol{X}$ and $\boldsymbol{X}^{(j)} = p_j \circ \tilde{p}_j(\boldsymbol{X}^{(j-1)})$ for each $j \in \{1, \ldots, d-1\}$.

Finally, $n$ columns from $\boldsymbol{X}^{(d-1)}$ are selected at random. Let $\tilde{\boldsymbol{X}}$ be the resulting matrix, and let $\tilde{\boldsymbol{X}}_i = (\tilde{\boldsymbol{X}}_{1i}, \ldots, \tilde{\boldsymbol{X}}_{di})$ for every integer $i \in \{1, \ldots, n\}$; then $\{\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n\}$ is a sample from the specified model.

# Chapter 4

# Inference for aggregation-tree models

To build an appropriate model for $(X_1, \ldots, X_d)$, one must select (i) a tree structure for the model; (ii) a collection of marginal distributions $F_1, \ldots, F_d$ for $X_1, \ldots, X_d$; and (iii) a collection of $d - 1$ copulas to join these variables in the construction process. Procedures for doing so are reviewed here in turn.

## 4.1 Tree structure selection

Although the number of possible tree structures for a model with $d$ variables is finite, the problem of choosing a tree structure for the model cannot be realistically solved by examining all possible structures. Indeed, let $N_d$ be the number all possible tree structures for a model with $d$ variables. The value of $N_d$ can be obtained as follows from a recursive formula. First, it is clear that $N_1 = 1$; then, for each integer $k \geq 2$,

$$N_k = \frac{1}{2} \sum_{i=1}^{k-1} \binom{d}{i} N_i \times N_{k-i}.$$

The number $N_d$ becomes unreasonably large even for relatively small values of $d$, e.g., for $d = 10$, $N_d > 3 \times 10^8$. Thus, if an explicit aggregation order is not chosen, one requires a systematic approach to select a tree structure for these models. Côté & Genest

(2015) suggest a method to select a tree structure for the model using classical hierarchical clustering methods. Specifically, they propose using the value of a measure of the absolute dependence between these variables as the dissimilarity metric. The indices of the variables combined at each step of the clustering process provide the tree structure $\mathcal{T}_1, \ldots, \mathcal{T}_d$.

Intuitively, this provides a sequence of clusterings where, at each step, a pair of variables is combined and the variables with the highest dependence level are combined first. It is suggested to use pseudo-metrics for the dissimilarity measures, as hierarchical clustering tends to perform better with such a dissimilarity function. In particular, the measures

$$D_r(X,Y) = \sqrt{1 - r^2(X,Y)},$$

$$D_\rho(X,Y) = \sqrt{1 - \rho^2(X,Y)} \quad \text{and} \quad D_\tau(X,Y) = \sqrt{1 - \tau^2(X,Y)},$$

where $r$ is the Pearson correlation, $\rho$ is the Spearman rank correlation, and $\tau$ is the Kendall rank correlation, are all pseudo-metrics.

## 4.2 Selection of copulas

To select the $d-1$ copulas that are used to model the dependence structure between the pairs of variables which are aggregated at each step of the construction of the model, Côté & Genest (2015) suggest a rank-based approach to help guide the choice of copulas for the model.

Let $B_{1i} = A_{\ell_i i}$ and $B_{2i} = A_{k_i i}$ be the index sets from the tree structure such that $M_{B_{1i}}$ and $M_{B_{2i}}$ are combined at the $i$th step of the tree construction. One can start by examining the rank plot of $(M_{B_{1i}}, M_{B_{2i}})$ visually to look for certain properties of the copula of $(M_{B_{1i}}, M_{B_{2i}})$. This way, one can select or filter out potential choices of copula families for $C_{B_{1i}B_{2i}}$, e.g., if the rank plot of $(M_{B_{1i}}, M_{B_{2i}})$ displays high levels of upper-tail dependence, a copula that will reflect this feature of the dependence structure, such as a Gum-

bel copula, may be appropriate. Inference procedures may also be used to rule out copula families. For instance, if we obtain a significant result from a test of asymmetry, such as the one proposed by Genest et al. (2012), this indicates that the copula of $(M_{B_{1i}}, M_{B_{2i}})$ is asymmetric and thus we may discard symmetric copulas as suitable choices for $C_{B_{1i}B_{2i}}$. Once a copula for $C_{B_{1i}B_{2i}}$ has been selected, one can use goodness-of-fit tests for bivariate copula models such as the one presented in Genest et al. (2011) to validate our choice.

## 4.3 Inference and model validation

Once a tree structure, a collection of copulas, and marginal distributions have been selected, Côté & Genest (2015) suggest a recursive inference approach to validate an aggregation-tree model. It consists of a recursive application of the copula comparison test from Rémillard & Scaillet (2009). These authors present a series of inference tests based on Cramér–von Mises type statistics to test for the equality of the copulas of two random samples.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m \in \mathbb{R}^d$ and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \in \mathbb{R}^d$ be independent random samples with distributions $F_1, \ldots, F_d$ and $G_1, \ldots, G_d$ and copulas $C$ and $D$, respectively. The empirical copula processes $\mathbb{C}_m = \sqrt{m}\,(\hat{C}_m - C)$ and $\mathbb{D}_n = \sqrt{n}\,(\hat{D}_n - D)$ both converge weakly to continuous centered Gaussian processes, i.e., $\mathbb{C}_m \rightsquigarrow \mathbb{C}$ and $\mathbb{D}_n \rightsquigarrow \mathbb{D}$. Now suppose that $n/(n+m) \to \lambda \in (0,1)$ as $\min(m,n) \to \infty$. We get, as $\min(m,n) \to \infty$,

$$\sqrt{\frac{n}{n+m}}\,\mathbb{C}_m - \sqrt{\frac{m}{n+m}}\,\mathbb{D}_n \rightsquigarrow \sqrt{1-\lambda}\,\mathbb{C} - \sqrt{\lambda}\,\mathbb{D}.$$

Under the null hypothesis $\mathcal{H}_0 : C = D$, for

$$\mathbb{E}_{n,m} = \sqrt{\frac{nm}{n+m}}\,(\hat{C}_m - \hat{D}_n) = \sqrt{\frac{n}{n+m}}\,\mathbb{C}_m - \sqrt{\frac{m}{n+m}}\,\mathbb{D}_n,$$

we get $\mathbb{E}_{n,m} \rightsquigarrow \sqrt{1-\lambda}\,\mathbb{C} - \sqrt{\lambda}\mathbb{D}$. With this asymptotic distribution for $\mathbb{E}_{n,m}$, Rémillard & Scaillet (2009) suggest the following test statistic to test $\mathcal{H}_0 : C = D$ vs $\mathcal{H}_1 : C \neq D$:

$$T_{m,n} = \frac{nm}{n+m} \int_{[0,1]^d} \{\hat{C}_m(u) - \hat{D}_n(u)\}^2 \, du.$$

Large values of this statistic provide evidence against the null hypothesis and Rémillard & Scaillet (2009) describe methods to compute the $p$-value.

Ways of computing $p$-values for this test have been implemented in the R `TwoCop` package (Rémillard & Plante 2012). For each $i \in \{1, \ldots, d-1\}$, let $D_i$ be the copula that is obtained for $\boldsymbol{X}_{A_{1i}} = (X_j : j \in A_{1i})$ from the selected copulas and tree structure of the model. Moreover, let $D_i^*$ be the true underlying copula of $(X_j : j \in A_{1i})$.

The procedure consists of comparing each $D_i$ with the corresponding $D_i^*$. We first perform the comparison test on $D_{d-1}$ and if the null $\mathcal{H}_0 : D_{d-1} = D_{d-1}^*$ fails to be rejected, the same test is then performed for $D_{d-2}$. The process is repeated until $\mathcal{H}_0 : D_i = D_i^*$ is rejected for some $i$ or we don't obtain a significant result for $\mathcal{H}_0 : D_1 = D_1^*$ (while adjusting for multiple testing), which does not indicate that there was an error with the tree structure or the copulas chosen for our model.

To perform these tests, one must have a sample from each copula $D_i$ and $D_i^*$ to compute their empirical copula estimators for the test statistic. For the $D_i^*$'s, the original data are used. For the $D_i$ copulas, one can use a sample generated with the Iman–Conover algorithm from Chapter 3.

This inference procedure serves to indirectly verify that the conditional independence assumption is satisfied. A significant test result indicates that there is an issue with the model, which could be caused by the conditional independence assumptions not being satisfied, among many things. To formally verify that the conditional independence is satisfied, one may also use multiple conditional independence tests to verify all the conditional independence assumptions for the leaf descendants of each branching node.

However, for large tree structures this may be unfeasible, especially when accounting for multiple testing. Thus, one may need to resort to this indirect approach.

# Chapter 5

# Aggregation-tree models in the extreme-value setting

As we wish to explore the application of aggregation-tree models in extreme-value theory, we are particularly interested in the properties of aggregation-tree models when we work with maxima, i.e., where the aggregation function is $a(x, y) = \max(x, y)$, and we only use extreme-value copulas to join the aggregated variables. Thus, we must first introduce some preliminary results about extreme-value copulas.

## 5.1 Extreme-value copulas and max-stable distributions

Consider a sample $(X_{11}, \ldots, X_{1d}), \ldots, (X_{n1}, \ldots, X_{nd})$ of size $n$ from a continuous $d$-variate distribution $F$ with marginals $F_1, \ldots, F_d$. Let $C$ denote the unique underlying copula. Next, consider the vector of component-wise maxima

$$\mathrm{M}_n = (\mathrm{M}_{n1}, \ldots, \mathrm{M}_{nd}),$$

where for each $j \in \{1, \ldots, d\}$, $\mathrm{M}_{nj} = \max(X_{1j}, \ldots, X_{nj})$. The distribution function of each $M_{nj}$ is given by $F_j^n$ and so the copula $C_n$ for $\mathrm{M}_n$ is given, for all $u_1, \ldots, u_d \in (0,1)$, by

$$\begin{aligned} C_n(u_1, \ldots, u_d) &= \mathbb{P}\{F_1^n(\mathrm{M}_{n1}) \leq u_1, \ldots, F_d^n(\mathrm{M}_{nd}) \leq u_d\} \\ &= \left[\mathbb{P}\{F_1(X_1) \leq u_1^{1/n}, \ldots, F_d(X_d) \leq u_d^{1/n}\}\right]^n = \left\{C(u_1^{1/n}, \ldots, u_d^{1/n})\right\}^n. \end{aligned}$$

**Definition 5.1.1** (Extreme-value copula)**.** A copula $C : [0,1]^d \to [0,1]$ is called an extreme-value copula if there exists a copula $C_*$ such that, for all $u_1, \ldots, u_d \in [0,1]$,

$$\lim_{n\to\infty} C_*^n(u_1^{1/n}, \ldots, u_d^{1/n}) = C(u_1, \ldots, u_d).$$

The copula $C_*$ is said to be in the domain of attraction of $C$.

Alternatively, there is a similar property called max-stability that helps to characterize the class of extreme-value copulas.

**Definition 5.1.2** (Max-stability)**.** A copula $C : [0,1]^d \to [0,1]$ is called a max-stable copula if and only if, for all $u_1, \ldots, u_d \in [0,1]$ and for all $n \in \mathbb{N}$,

$$\left\{C(u_1^{1/n}, \ldots, u_d^{1/n})\right\}^n = C(u_1, \ldots, u_d).$$

Clearly, a max-stable copula is an extreme-value copula and it belongs to its own domain of attraction. Interestingly, max-stability is actually also a necessary condition for a copula to be extreme-value.

**Theorem 5.1.1.** *A copula is an extreme-value copula if and only if it is max-stable.*

*Proof.* Given what has already been said, it is only necessary to show that every extreme-value copula is also max-stable. Let $C$ be an extreme-value copula. Then there exists some copula $C_*$ such that, for all $u_1, \ldots, u_d \in (0,1)$,

$$\lim_{m\to\infty} C_*^m(u_1^{1/m}, \ldots, u_d^{1/m}) = C(u_1, \ldots, u_d).$$

Now fix $n \in \mathbb{N}$. Then for all $u_1, \ldots, u_d \in (0,1)$, one has

$$C(u_1, \ldots, u_d) = \lim_{k\to\infty} \left\{C_*(u_1^{1/(kn)}, \ldots, u_d^{1/(kn)})\right\}^{kn}$$
$$= \left[\lim_{k\to\infty} C_*^k\{(u_1^{1/n})^{1/k}, \ldots, (u_d^{1/n})^{1/k}\}\right]^n = \left\{C(u_1^{1/n}, \ldots, u_d^{1/n})\right\}^n,$$

where the map $(u_1, \ldots u_d) \mapsto C_*^k(u_1^{1/k}, \ldots, u_d^{1/k})$ defines the copula of the vector of component-wise maxima of a sample of size $k$ from the copula $C_*$. Thus, $C$ is max-stable. □

Extreme-value distributions and their properties are one of the main subjects of analysis in extreme-value theory. Extreme-value distributions represent the limiting distribution of normalized sample maxima. A result by Fisher & Tippett (1928), refined by Gnedenko (1943), shows that in dimension $d = 1$, there exists exactly three possible types for such limiting distributions; see Proposition 0.3 in Resnick (1987).

Before stating this result, note that two distribution functions $F$ and $G$ are said to be of the same type if there exists $a \in (0, \infty)$ and $b \in \mathbb{R}$ such that, for all $x \in \mathbb{R}$, $F(x) = G(ax+b)$.

**Theorem 5.1.2** (Gnedenko). *Let $X_1, X_2, \ldots$ be an i.i.d sequence of random variables with distribution function $F$ and let $a_1, a_2, \ldots \in (0, \infty)$ and $b_1, b_2, \ldots \in \mathbb{R}$ be sequences of scalars such that, for $\mathrm{M}_n = \max(X_1, \ldots, X_n)$ and all $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\mathrm{M}_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x),$$

*where $G$ is a non-degenerate distribution function. Then $G$ is of one of the three following types:*

*Fréchet: $\Phi_\alpha(x) = \exp(-x^{-\alpha})$ for $x > 0$ and $\Phi_\alpha(x) = 0$ for $x \leq 0$ with $\alpha \in (0, \infty)$.*

*Weibull: $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$ for $x < 0$ and $\Psi_\alpha(x) = 1$ for $x \geq 0$ with $\alpha \in (0, \infty)$.*

*Gumbel: $\zeta(x) = \exp(-e^{-x})$ for all $x \in \mathbb{R}$.*

The members of the parametric classes $\Phi_\alpha$, $\Psi_\alpha$ and $\zeta$ are called the extreme-value distributions and a random variable $Y \in \mathbb{R}$ is called extreme-value if its distribution belongs

to one of these three families. Univariate extreme-value distributions are used in tandem with extreme-value copulas to define multivariate extreme-value distributions.

**Definition 5.1.3** (Multivariate extreme-value distributions)**.** A continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ has a multivariate extreme-value distribution (MEVD) $F$ if and only if its underlying copula is extreme-value and the univariate margins $F_1, \ldots, F_d$ of its components $X_1, \ldots, X_d$ are all extreme-value distributions.

We will work, for the most part, with multivariate extreme-value distributions with unit Fréchet margins, i.e., with cumulative distribution function given, for all $x \in (0, \infty)$, by $\Phi_1(x) = \exp(-1/x)$. These distributions are particularly easy to work with, and they possess some important properties for our analysis. Proposition $5.11$ from Resnick (1987) provides a useful characterization of multivariate extreme-value distributions with unit Fréchet margins.

**Theorem 5.1.3.** *The following statements are equivalent:*

*(i)* $\boldsymbol{X} = (X_1, \ldots, X_d)$ *is a* $d$*-dimensional extreme-value random vector with unit Fréchet margins.*

*(ii) There exists a finite measure* $\sigma$ *on the unit simplex*

$$S = \{\boldsymbol{a} \in [0,1]^d : ||\boldsymbol{a}||_1 = a_1 + \cdots + a_d = 1\}$$

*such that for all* $i \in \{1, \ldots, d\}$, $\int_S a_i d\sigma(a) = 1$ *and, for all* $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \exp\left\{-\int_S \max\left(\frac{a_1}{x_1}, \ldots, \frac{a_d}{x_d}\right) d\sigma(a)\right\}.$$

*The measure* $\sigma$ *is called a spectral measure.*

*(iii) There exist $f_1, \ldots, f_d$ non-negative Lebesgue measurable functions on $[0,1]$ such that for all $i \in \{1, \ldots, d\}$, $\int_0^1 f_i(x)dx = 1$ and, for all $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \exp\left[-\int_0^1 \max\left\{\frac{f_1(u)}{x_1}, \ldots, \frac{f_d(u)}{x_d}\right\} du\right],$$

*where $\boldsymbol{X} \leq \boldsymbol{x}$ denotes element-wise inequality.*

As a consequence of this result, if a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ has an extreme-value distribution with unit Fréchet margins, then $\mathrm{M} = \max(X_1, \ldots, X_d)$ is a scaled Fréchet random variable. This follows as one has, for all $x \in [0, \infty)$,

$$\mathbb{P}(\mathrm{M} \leq x) = \exp\left[-\int_0^1 \max\left\{\frac{f_1(u)}{x}, \ldots, \frac{f_d(u)}{x}\right\} du\right] = \exp\left(-s/x\right),$$

where $s = \int_0^1 \max\left\{f_1(u), \ldots, f_d(u)\right\} du$. These characterizations will prove useful to establish some results about the joint distribution obtained from an aggregation-tree model that uses only extreme-value copulas and unit Fréchet margins. There also exist important characterizations for extreme-value copulas from Pickands (1981).

**Theorem 5.1.4** (Pickands)**.** *Let $C : [0,1]^d \to [0,1]$ be an extreme-value copula. Then there exists a spectral measure $\sigma$ on the unit simplex $S$ such that, for all $\boldsymbol{u} = (u_1, \ldots, u_d) \in (0,1)^d$,*

$$C(u_1, \ldots, u_d) = \exp\left[\left\{\sum_{i=1}^d \ln(u_i)\right\} \times B\left\{\frac{\ln(u_1)}{\sum_{i=1}^d \ln(u_i)}, \ldots, \frac{\ln(u_{d-1})}{\sum_{i=1}^d \ln(u_i)}\right\}\right],$$

*where $B : \mathbb{R}^{d-1} \to [0, \infty)$ is given by*

$$B(u_1, \ldots, u_{d-1}) = \int_S \max\left\{x_1 u_1, \ldots, x_{d-1} u_{d-1}, x_d\left(1 - \sum_{i=1}^{d-1} u_i\right)\right\} d\sigma(x).$$

The map $B$ is called the Pickands dependence function and it possesses the following properties:

a) For all $(u_1, \ldots, u_{d-1}) \in \{\boldsymbol{x} \in [0,1]^{d-1} : x_1 + \cdots + x_{d-1} \leq 1\}$, one has

$$\max\left(u_1, \ldots, u_{d-1}, 1 - \sum_{i=1}^{d-1} u_i\right) \leq B(u_1, \ldots, u_{d-1}) \leq 1.$$

b) $B$ is convex on its domain.

c) $B(e_i) = 1$ for any $i \in \{1, \ldots, d-1\}$, where $e_i$ is the $i$th vector from the standard basis of $\mathbb{R}^{d-1}$, i.e., $e_{ij} = 1$ if $i = j$ and $e_{ij} = 0$ otherwise.

The bivariate version of this result is of particular interest for aggregation-tree models since the latter rely on bivariate copulas. Consider the map $A : [0,1] \to \mathbb{R}$ given by $A(t) = B(1-t)$ for all $t \in [0,1]$, which is also called the Pickands dependence function. Then any bivariate extreme-value copula $C$ can be written, for all $u, v \in (0,1)$, as

$$C(u,v) = \exp\left[\ln(uv) A\left\{\frac{\ln(v)}{\ln(uv)}\right\}\right].$$

Thus, the Pickands dependence function characterizes extreme-value copulas. For instance, the Pickands dependence function of the bivariate Gumbel copula with parameter $\theta \in [1, \infty)$ is given, for all $t \in [0,1]$, by $A_\theta(t) = \{t^\theta + (1-t)^\theta\}^{1/\theta}$.

Pickands dependence functions can be used to compute various quantities associated with the dependence of random vectors whose copula is extreme-value. Section 6.4 in Jaworski et al. (2010) provides formulas relating the Pickands dependence function to various dependence coefficients, such as Kendall's $\tau$, Spearman's $\rho$, and the upper tail dependence coefficient $\lambda_U$, viz.

$$\tau(X,Y) = \int_0^1 \frac{t(1-t)}{A(t)}\, dA'(t),$$

$$\rho(X,Y) = -3 + 12 \int_0^1 \frac{1}{(1+A(t))^2}\, dt,$$

$$\lambda_U(X,Y) = 2\{1 - A(1/2)\}.$$

## 5.2 Tree structure selection for extreme-value models

Following the hierarchical clustering techniques to choose a tree structure for a model suggested by Côté & Genest (2015), we want to find a pseudo-metric based on a dependence measure that would be suitable in the extreme-value setting. A natural choice for this setting is to create a dissimilarity measure based on one of the tail dependence coefficients.

Murphy (2018) presents a clustering analysis of 23 rainfall gauging stations in the province of Québec where she used a dissimilarity measure based on the upper-tail dependence coefficient as defined in Section 1.3. This dissimilarity measure is the $F$-madogram from Cooley et al. (2006), whose definition is recalled below.

**Definition 5.2.1** ($F$-madogram)**.** Let $(X, Y)$ be a random pair with joint distribution function $F$ and marginal distribution functions $F_X$ and $F_Y$. The corresponding $F$-madogram is defined by

$$M(X,Y) = \frac{1}{2}\,\mathrm{E}\big\{|F_X(X) - F_Y(Y)]|\big\}.$$

The $F$-madogram possesses some desirable properties for the purpose of model selection, viz.

(a) $M(X,Y)$ is a pseudo-metric as it is symmetric $M(Y,X) = M(X,Y)$. It is also positive with $M(X,X) = 0$ and verifies a sort of triangle inequality given that for any continuous random variable $Z$ with cumulative distribution function $F_Z$,

$$\begin{aligned}\mathrm{E}\{|F_X(X) - F_Y(Y)]|\} &= \mathrm{E}[|\{F_X(X) - F_Z(Z)\} - \{F_Z(Z) - F_Y(Y))\}|] \\ &\leq \mathrm{E}\{|F_X(X) - F_Z(Z)|\} + \mathrm{E}\{|F_Z(Z) - F_Y(Y))|\},\end{aligned}$$

which implies, $M(X,Y) \leq M(X,Z) + M(Z,Y)$.

(b) Given that $F_X(X)$ and $F_Y(Y)$ are both standard uniform random variables,

$$M(Y,X) \leq \frac{1}{2}\left(\mathrm{E}\{|F_X(X)|\} + \mathrm{E}\{|F_Y(Y))|\}\right) = \frac{1}{2}.$$

Therefore, $0 \leq M(X,Y) \leq 1/2$.

(c) If $X$ and $Y$ are independent, $M(Y,X) = 1/6$. This follows as

$$M(Y,X) = \frac{1}{2}\int_0^1 \int_0^1 |x_1 - x_2| dx_1 dx_2 = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}.$$

However, one of the most important features of the $F$-madogram for selection methods in the extreme-value setting is its relation to the upper-tail dependence coefficient. This identity is presented in Murphy (2018).

**Theorem 5.2.1.** *Let $(X,Y)$ be a random pair with joint distribution function $F$ and an extreme-value copula, and let $\lambda_U(X,Y)$ be the upper tail dependence coefficient for $(X,Y)$. Then*

$$M(X,Y) = \frac{1 - \lambda_U(X,Y)}{2\{3 - \lambda_U(X,Y)\}}.$$

The proof for this result is presented in Murphy (2018). This result indicates why $F$-madograms are appropriate for our selection method: $M(X,Y)$ is a decreasing function of $\lambda_U(X,Y)$ and is maximized when $\lambda_U(X,Y) = 0$, i.e., when $X$ and $Y$ are upper-tail independent. Furthermore, $M(X,Y) = 0$ when $\lambda_U(X,Y) = 1$, i.e., when $X$ and $Y$ are perfectly upper-tail dependent.

Using $M(X,Y)$ as a dissimilarity measure for hierarchical clustering will generate an aggregation-tree where the most tail dependent random variables will be joined first. We require a method to estimate $F$-madograms to be able to employ it in our tree structure selection method. One approach is to estimate the $F$-madogram directly.

A rank-based estimator for the $F$-madogram was suggested by Naveau et al. (2009). Given a random sample $(X_1,Y_1)\ldots,(X_n,Y_n)$ from the pair $(X,Y)$, let $\hat{F}_{X,n}$ and $\hat{F}_{Y,n}$ denote the empirical distribution functions of $X$ and $Y$, respectively. The suggested estimator for the $F$-madogram of $X$ and $Y$ is given by

$$\hat{M}(X,Y) = \sum_{i=1}^{n} \frac{1}{2n} |\hat{F}_{X,n}(X_i) - \hat{F}_{Y,n}(Y_i)| = \sum_{i=1}^{n} \frac{1}{2n^2} |R_{i1} - R_{i2}|,$$

where $R_{i1}$ is the rank of $X_i$ among $X_1, \ldots, X_n$ and $R_{i2}$ is the rank of $Y_i$ among $Y_1, \ldots, Y_n$. Alternatively, under the assumption that the copula that joins the variables is extreme-value, we may simply estimate the upper tail dependence coefficient instead and use the previous formula to obtain an estimate for $M(X, Y)$.

Note that this measure is essentially the same as Spearman's footrule, considered by Genest et al. (2010). The latter is given by

$$\hat{S}(X, Y) = 1 - \frac{1}{n^2 - 1} \sum_{i=1}^{n} |R_{i1} - R_{i2}|,$$

and its asymptotic behavior was considered in that paper.

# Chapter 6

# Data illustration

We proceed to illustrate the process of fitting an extreme-value aggregation-tree model using a dataset collected from Québec hydrometric stations. For this example, we will examine five particular stations near the island of Montréal (or Kawenote Teiontiakon) which monitor the flow on the St. Lawrence (or Magtogoek) river and on some of its tributaries. The call codes for these five stations are given in Table 6.1, and their geographical locations are indicated on the map in Figure 6.1.

The dataset consists of mean monthly discharge $(\mathrm{m}^3/\mathrm{s})$ measurements collected from each of the stations for every month from January $2000$ to December $2019$, which amounts to $240$ observations. Let $X_1, \ldots, X_5$ be random variables representing the readings at each of the five stations. For each $i \in \{1, \ldots, 240\}$ and $j \in \{1, \ldots, 5\}$, let $X_{ij}$ be the $i$th observation from the $j$th station. As the models and inference techniques were designed to work with mutually independent and identically distributed data, we must first assess the presence of seasonality and other trends in the time series data from these stations.

The time series of collected data from each of these stations is presented in Figure 6.2. From the context of the study, it is safe to assume that there is a seasonal effect and perhaps a trend (possibly from the effects of climate change) in the data for all five stations, so we use nonparametric inference tests to check for the presence of both.

**Table 6.1:** Station code and river being monitored by each of the hydrometric stations used in the dataset.

| Variable | River | Station Code |
|---|---|---|
| $X_1$ | Rivière des Mille Îles | 02OA003 |
| $X_2$ | Saint-Laurent | 02OA016 |
| $X_3$ | Rouge | 02LC029 |
| $X_4$ | Rivière de l'Achigan | 02OB037 |
| $X_5$ | Châteauguay | 02OA054 |

Using a seasonality test based on the Kruskal–Wallis test, we obtain $p$-values very close to zero ($p < 10^{-9}$) for the data for all five stations. Thus, we use the STL decomposition as presented in Cleveland et al. (1990) to remove the seasonality and trend from the station data, viz.

$$X_{ij} = S_{ij} + T_{ij} + R_{ij}.$$

For each station's time series, the seasonal effect $S_{ij}$ is computed by taking the average of all the observations of the station recorded in the month that $X_{ij}$ was recorded. The trend effect $T_{ij}$, if any, is estimated using a LOESS window of size $12$ months, corresponding to a period of one year.

We attempt to fit an aggregation-tree model for the residuals $R_{ij}$ of the time series models for $X_1, \ldots, X_5$ using maxima and extreme-value copulas for the aggregation. The use of models on the residuals is justified by the fact that under certain regularity conditions, the empirical copula process obtained from the model residuals behaves as if the innovations were observed, as shown by Rémillard (2017). We will also attempt to fit a $5$-dimensional extreme-value copula model to this dataset. This data analysis will serve to illustrate the richness and flexibility of aggregation-tree models and how they compare to classical copula models.


**Figure 6.1:** Map showing the physical location of the five hydrometric stations listed in Table 6.1.

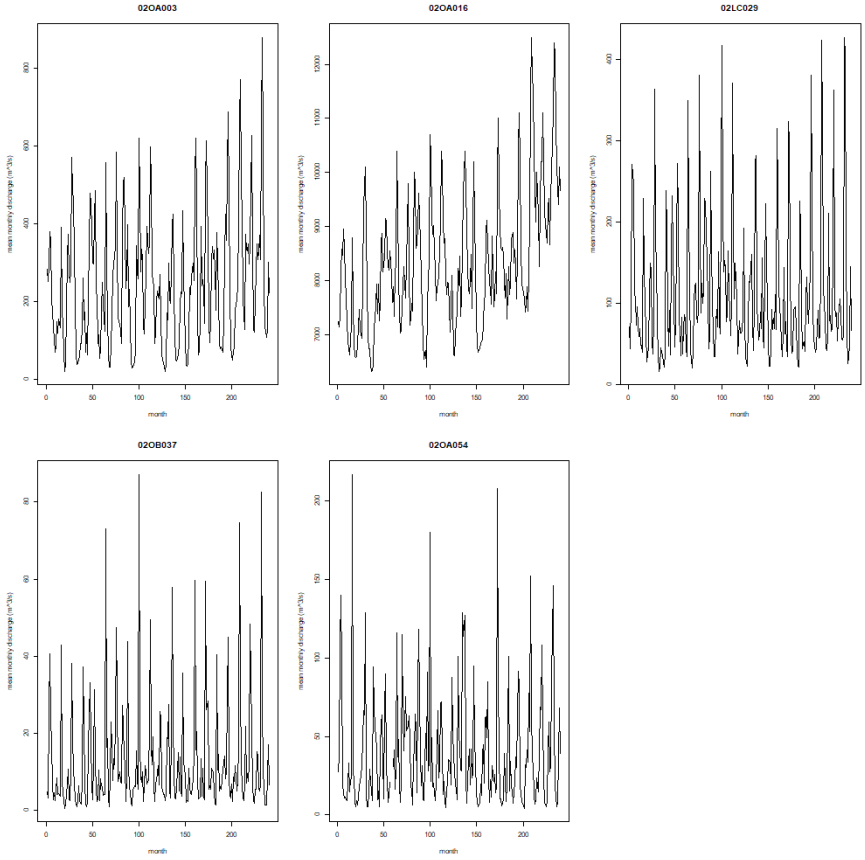


**Figure 6.2:** Time series of hydrometric data recorded at each station.

## 6.1 Tree structure selection

Before we select the tree structure for the model, we transform the data for each station to ensure that the margins are unit Fréchet. If $F_j$ is the distribution function of $X_j$ for each $j \in \{1, \ldots, 5\}$, then $Y_j = -1/\ln\{F_j(X_j)\}$ is unit Fréchet distributed. However, because $F_j$ is unknown, it must be estimated. Therefore, we will proceed with the transformed data defined, for all $i \in \{1, \ldots, 240\}$ and $j \in \{1, \ldots, 5\}$, by

$$\tilde{X}_{ij} = \frac{-1}{\ln\{\hat{F}_j(X_{ij})\}},$$

where for each $j \in \{1, \ldots, 5\}$, $\hat{F}_j$ is the empirical distribution function obtained from the univariate sample $X_{1j}, \ldots, X_{nj}$.

Once the data have been transformed to have approximately unit Fréchet margins, we use the hierarchical clustering approach described by Côté & Genest (2015) with the $F$-madogram dissimilarity measure. Using the dissimilarity matrix $D_{ij} = \hat{M}(X_i, X_j)$, where $\hat{M}(X_i, X_j)$ is the rank-based estimate of the $F$-madogram from Naveau et al. (2009), we perform hierarchical clustering of the stations with single linkage.

The dendrogram displayed in Figure 6.3 suggests the tree structure $\mathcal{T}$ with $\mathcal{T}_2 = \{\{1,3\},\{2\},\{4\},\{5\}\}$, $\mathcal{T}_3 = \{\{1,2,3\},\{4\},\{5\}\}$, and $\mathcal{T}_4 = \{\{1,2,3,4\},\{5\}\}$.

## 6.2 Copula selection

Based on the hierarchical clustering technique, we set

$$M_{A_{11}} = \max(X_1, X_3), \quad M_{A_{12}} = \max(X_1, X_2, X_3), \quad M_{A_{13}} = \max(X_1, X_2, X_3, X_4).$$

We must select four copulas, namely $C_{\{1\}\{3\}}$ for $(X_1, X_3)$, $C_{\{1,3\}\{2\}}$ for $(M_{A_{11}}, X_2)$, $C_{\{1,2,3\}\{4\}}$ for $(M_{A_{12}}, X_4)$, and $C_{\{1,2,3,4\}\{5\}}$ for $(M_{A_{13}}, X_5)$.

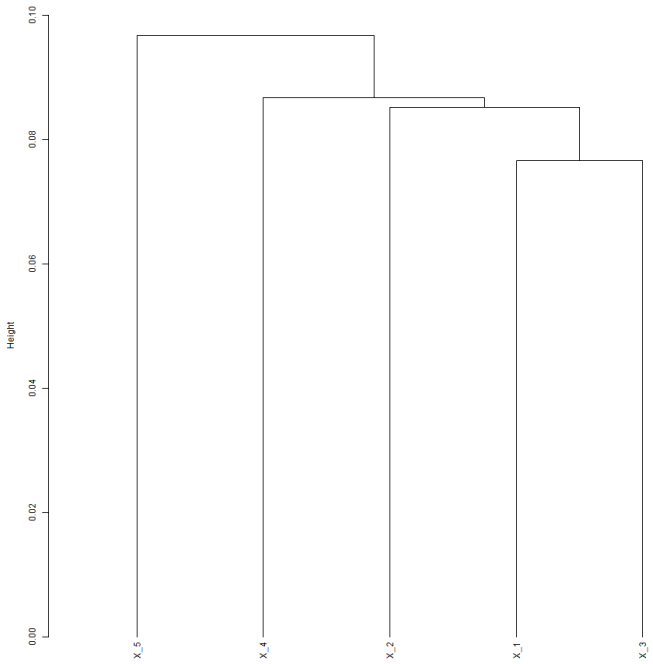


**Figure 6.3:** Dendrogram of the hierarchical clustering of the hydrometric station data using the $F$-madogram as the dissimilarity measure.

Initially, examining the rank plots of these variables may help to spot certain features of the copulas of these pairs. This helps to rule out certain copula families and select which properties should be verified using statistical tests.

The rank plots, provided in Figure 6.4, seem to suggest that there is some level of upper tail dependence in each of these pairs. In contrast, the lower tail dependence seems to be rather weak in all cases, in particular in the rank plot of the pair $(M_{A_1}, X_2)$. This rules out certain copula families such as the Clayton model or any copula family without upper tail dependence for all four cases.

The rank plots of the pairs $(X_1, X_3)$ and $(M_{A_{12}}, X_4)$ suggest that the copulas for these pairs are symmetric. The symmetry for the copulas for the other pairs is less obvious from the rank plots. We use tests for independence, extremeness, and symmetry to guide our choice for the copulas as well.

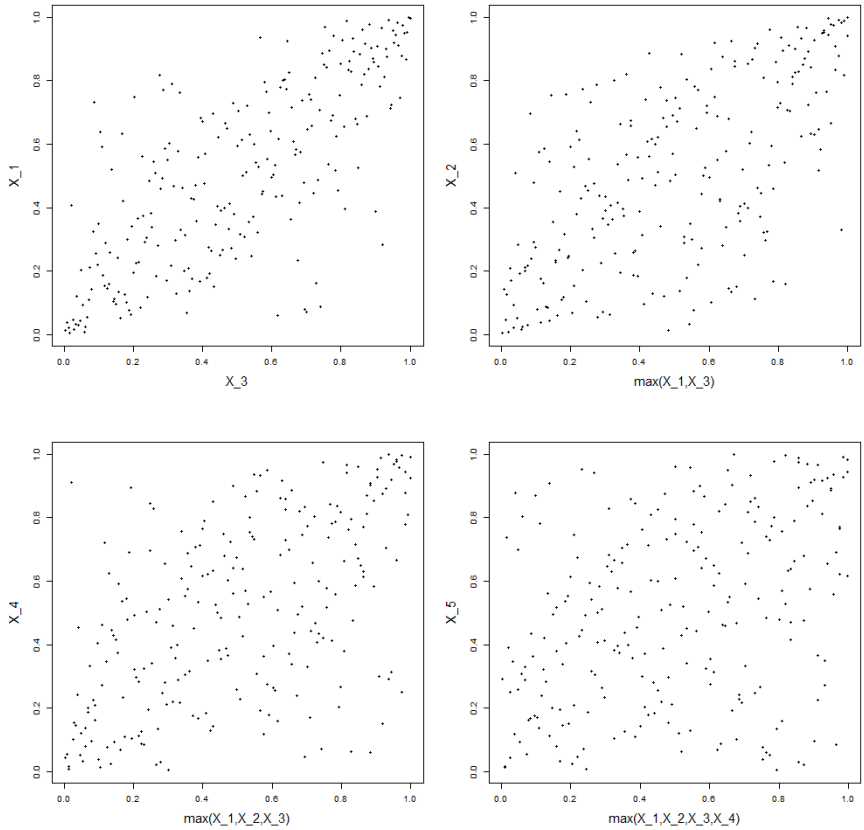


**Figure 6.4:** Rank plots of the modified sample used to guide the choice of copulas.

We begin by testing the independence for all four pairs using the method implemented in the R package `copula` (Hofert et al. 2023). These tests are based on the work of Genest & Rémillard (2004). For each of the four pairs, the $p$-values for the test is near zero, which indicates that the independence copula is inadequate in all cases.

Moreover, as we want to use extreme-value copulas for this model, we use a test of extremeness for bivariate copulas from Ghoudi et al. (1998), further refined by Ben Ghorbal et al. (2009). The null hypothesis of this test is that the copula of the pair is extreme-value; therefore, large $p$-values (say, greater than 0.05) indicate that there is no evidence that the copula is not extreme-value.

The $p$-values for these statistical tests can be found in Table 6.3. For three of these pairs, the $p$-values for all the extreme-value tests are relatively high, the lowest $p$-value being 0.1994. So there is little evidence to suggest that an extreme-value copula is an

inadequate choice for any of these pairs. For the pair $(M_{A_{12}}, X_4)$, the $p$-value is 0.0519, which is close to, but still larger than, 0.05. For this reason, one can safely assume an extreme-value copula for the pair $(M_{A_{12}}, X_4)$ as well.

Proceeding under the assumption that the underlying copulas of all pairs are extreme-value, we then resort to a test of exchangeability for extreme-value copulas from Kojadinovic & Yan (2012) to determine whether we should choose symmetric copulas. We also apply an omnibus exchangeability test from Genest et al. (2012), which applies to any bivariate copula, whether it is extreme-value of not. As with the test for extremeness, large $p$-values are indicative of lack of evidence against the copula being exchangeable.

The $p$-values for both these statistical tests are reported in Table 6.4. We see that with two pairs, $(X_1, X_3)$ and $(M_{A_{13}}, X_4)$, the $p$-values for the two exchangeability tests are high. The $p$-value for the EV-exchangeability test for the pair $(M_{A_{12}}, X_4)$, $p = 0.08242$, is small but still larger than 0.05. The $p$-value for the other exchangeability test for the copula of $(M_{A_{12}}, X_4)$ is larger ($p = 0.1344$). This suggests that we may use symmetric copulas for these pairs.

The $p$-value for the EV-exchangeability test for the pair $(M_{A_{11}}, X_2)$ was quite small ($p = 0.0085$). However, the $p$-value for the other exchangeability test is much larger, $p = 0.2493$. Upon visual examination of the rank plot for this pair, we notice that the copula of this sample is slightly left skewed, so this result is not unexpected. However, given that this skewness appears very slight in the rank plot, it is possible that a symmetric copula could still be adequate for this pair. We proceed to fit symmetric extreme-value copulas for $C_{\{1\}\{3\}}, C_{\{1,2,3\}\{4\}}$ and $C_{\{1,2,3,4\}\{5\}}$, and we will attempt to fit both a symmetric extreme-value copula $C^{(1)}_{\{1,3\}\{2\}}$ and an asymmetric extreme-value copula $C^{(2)}_{\{1,3\}\{2\}}$ for the pair $(M_{A_{11}}, X_2)$.

As we require symmetric extreme-value copulas with non-zero upper-tail dependence coefficient, a natural choice is copulas from the Gumbel family. For the asymmetric copula $C^{(2)}_{\{1,3\}\{2\}}$, an asymmetric Gumbel copula obtained using Khoudraji's device would be a good option; see Khoudraji (1995) and Genest et al. (1998) for a description of this device.

We will use Gumbel copulas exclusively for the symmetric copulas $C_{\{1\}\{3\}}$, $C^{(1)}_{\{1,3\}\{2\}}$, $C_{\{1,2,3\}\{4\}}$, and $C_{\{1,2,3,4\}\{5\}}$ for the model and an additional asymmetric Gumbel copula $C^{(2)}_{\{1,3\}\{2\}}$ to model the dependence between $M_{A_{11}}$ and $X_2$. To select an adequate parameter $\theta$ value for each of the Gumbel copulas, we use three approaches.

For the first technique, we estimate $\theta$ based on an estimator for the Pickands dependence function of the Gumbel copula. The Pickands dependence function of a Gumbel copula is given by $A_\theta(t) = \{t^\theta + (1-t)^\theta\}^{1/\theta}$ for all $t \in [0,1]$, and in particular $A_\theta(1/2) = 2^{1/\theta - 1}$. Thus, we may use the following, moment-type, estimator for the parameter

$$\hat{\theta} = \frac{\ln(2)}{\ln\{2\hat{A}(1/2)\}},$$

where $\hat{A}$ is an estimator for the Pickands dependence function of the copula for the sample. We use the rank-based estimator from Genest & Segers (2009) to estimate the Pickands dependence function for the copula of each pair.

For the second approach, we may obtain an estimate of the $\theta$ parameter by inverting Kendall's $\tau$. For a Gumbel copula with parameter $\theta$, Kendall's $\tau$ is given by $\tau = 1 - 1/\theta$. Given the standard estimator $\hat{\tau}$ for Kendall's $\tau$, we may thus use the following estimator for the parameter of a Gumbel copula:

$$\tilde{\theta} = \frac{1}{1-\hat{\tau}},$$

For the third approach, we use the maximum pseudo-likelihood method due to Genest et al. (1995) and presented in Section 1.4.2. The `fitCopula` function from the R package `copula` provides the estimates and corresponding confidence intervals for second and third approaches.

The three sets of parameter estimates for the Gumbel copulas are provided in Table 6.2, along with asymptotic 95% confidence intervals. Note that the parameter values across all estimation methods become smaller as one ascends the tree structure, a reflection of the fact that dependence is weaker higher up in the tree structure. This is in accordance

**Table 6.2:** Parameter estimates for the Gumbel copulas generated by each method with 95% asymptotic confidence intervals.

| Pair | Inverse Pickands | Inverse $\tau$ | Pseudo-likelihood |
|---|---|---|---|
| $(X_1, X_3)$ | 2.209 | 2.264 (1.933, 2.596) | 2.234 (1.954, 2.514) |
| $(M_{A_{11}}, X_2)$ | 1.809 | 1.843 (1.606, 2.080) | 1.842 (1.587, 2.0968) |
| $(M_{A_{12}}, X_4)$ | 1.624 | 1.664 (1.444, 1.884) | 1.664 (1.459, 1.827) |
| $(M_{A_{13}}, X_5)$ | 1.340 | 1.329 (1.145, 1.513) | 1.325 (1.196, 1.462) |

with the selection method for the tree structure, as it was designed to join variables with higher levels of upper-tail dependence early on.

We proceed with the parameter estimates from the inversion of Kendall's $\tau$ approach for the Gumbel copulas. For comparison purposes, we also chose to estimate the parameters of a Khoudraji–Gumbel copula by the method of maximum pseudo-likelihood. The pseudo-likelihood is maximized using a Nelder–Mead approach. The asymmetric Gumbel copula is obtained from Khoudraji's device combining an independence copula and a Gumbel copula with parameter $\theta$. This copula also possesses two additional shape parameters, $\alpha_1$ and $\alpha_2$.

The Khoudraji–Gumbel copula is defined, for all $(u, v) \in (0, 1)^2$, by

$$C_{KG}(u, v) = u^{(1-\alpha_1)} v^{(1-\alpha_2)} \exp\left[ - \left\{ |\ln(u^{\alpha_1})|^{\theta} + |\ln(v^{\alpha_2})|^{\theta} \right\}^{1/\theta} \right].$$

The estimated parameters for the methods are $\theta = 1.933 \pm 0.470$, $\alpha_1 = 0.900 \pm 0.263$, and $\alpha_2 = 1.000 \pm 0.318$.

## 6.3 Model validation

Once the copulas $C_{\{1\}\{3\}}, C^{(1)}_{\{1,3\}\{2\}}, C_{\{1,2,3\}\{4\}}, C_{\{1,2,3,4\}\{5\}}$ for the aggregation have been selected, we proceed with the inference procedure discussed in Chapter 4 to validate our choice of copulas.

**Table 6.3:** Observed $p$-values for comparison test and distance metric to select a Gumbel or asymmetric Gumbel copula for $C_{\{1,3\}\{2\}}$.

| Copula Model | Goodness-of-fit Test | Cross-Validation CIC |
|---|---|---|
| Khoudraji–Gumbel | 0.3731 | 70.09 |
| Gumbel | 0.0285 | 72.19 |

First, we must determine whether $C^{(1)}_{\{1,3\}\{2\}}$ or $C^{(2)}_{\{1,3\}\{2\}}$ is more suitable for $(M_{A_{11}}, X_2)$. To this end, we use both a goodness-of-fit test and model selection criteria for copula models. We use the goodness-of-fit test for extreme-value bivariate copulas from Genest et al. (2011), which is based on a Cramér–von Mises statistic to check whether the Khoudraji–Gumbel copula or Gumbel copula model fits the transformed residuals for the pair $(M_{A_{11}}, X_2)$.

Moreover, we use an AIC-like model selection criterion for copula models called the cross-validation copula information criterion (CIC) from Grønneberg & Hjort (2014), as implemented in the R package `copula`. This criterion is based on cross-validated log (pseudo) likelihood for the copula model. The model with the lowest CIC value should be selected.

The $p$-value for the goodness-of-fit test and CIC values for both these copula models can be found in Table 6.3. The goodness-of-fit test suggests that there is a lack of fit with the Gumbel copula $C^{(1)}_{\{1,3\}\{2\}}$ ($p$-value $p = 0.0285$). We also see that the CIC value for the Khoudraji–Gumbel copula model is smaller than the CIC for the regular Gumbel model. Therefore, the asymmetric Gumbel copula $C^{(2)}_{\{1,3\}\{2\}}$ appears to be a more suitable choice for $C_{\{1,3\}\{2\}}$.

We also applied the goodness-of-fit test from Genest at al. (2011) for the selected Gumbel copulas for $C_{\{1\}\{3\}}$, $C_{\{1,2,3\}\{4\}}$, and $C_{\{1,2,3,4\}\{5\}}$. The $p$-values for these goodness-of-fit tests are presented in Table 6.4. We see that all the $p$-values for these tests are larger than 0.05; thus there is no evidence that the copula models that we selected for the copulas of these pairs fit the data poorly.

**Table 6.4:** Observed $p$-values for extreme-value and exchangeability tests and for the goodness-of-fit of each of the selected copulas for the model.

| Pair | EV test | Exch test | EV Exch test | G.O.F test |
|---|---|---|---|---|
| $(X_1, X_3)$ | 0.1994 | 0.6109 | 0.9266 | 0.455 |
| $(M_{A_{11}}, X_2)$ | 0.4719 | 0.2493 | 0.0085 | 0.373 |
| $(M_{A_{12}}, X_4)$ | 0.0516 | 0.1344 | 0.0824 | 0.178 |
| $(M_{A_{13}}, X_5)$ | 0.9283 | 0.0495 | 0.1743 | 0.247 |

Having decided to use an asymmetric Gumbel copula for $C_{\{1,3\}\{2\}}$, we generate a sample of the same size as the Québec hydrometric dataset from a joint distribution with unit Fréchet margins and the same tree structure and copulas that were selected with the Iman–Conover algorithm from Chapter 3. We use this generated sample to validate the model by comparing the copula of each of the vectors $(X_1, \ldots, X_5)$, $(X_1, \ldots, X_4)$, $(X_1, X_2, X_3)$, and $(X_1, X_3)$ from the original sample with the copula from the corresponding subvector from the generated sample with the comparison test from Rémillard & Scaillet (2009).

The tests must be applied recursively: the copula for $(X_1, \ldots, X_5)$ is compared and then if the $p$-value for the test is larger than 0.05, the test is applied to the copula of $(X_1, \ldots, X_4)$, and so on. The $p$-values for the validation tests are also presented in Table 6.5. We find that the $p$-values for the four comparison tests are larger than 0.05. This indicates that none of the copulas which we chose stray far from the real underlying copula of each of the modeled vectors.

**Remark.** Note that these applications of the copula comparison tests are ad hoc. The parameters which were chosen for each of the copulas used depend on the data. Therefore, the sample generated by the Iman–Conover algorithm is not independent of the hydrometric data, which is required to use the copula comparison test. Moreover, although $p$-values for the copula comparison tests are larger than 0.05, the comparison test for the

copula of $(X_1, \ldots, X_5)$ is still quite close to 0.05, so it is doubtful that one would still fail to reject the null if some adjustment to account for multiple testing had been used.

**Table 6.5:** $p$-values for the recursive validation approach.

| Vector | $(X_1, \ldots, X_5)$ | $(X_1, \ldots, X_4)$ | $(X_1, X_2, X_3)$ | $(X_1, X_3)$ |
|---|---|---|---|---|
| $p$-value | 0.08 | 0.27 | 0.24 | 0.89 |

Thus, we obtain a multivariate model for the joint transformed data of five hydrometric stations in the province where we used symmetric and asymmetric Gumbel copulas to model the dependence between the aggregated margins.

a) We transformed the original hydrometric data into pseudo-Fréchet data using probability integral transform. A tree structure was then selected for the aggregation-tree model using a hierarchical clustering technique based on the $F$-madogram for the transformed data of each station.

b) Gumbel copulas were used to model the dependence structure between $(X_1, X_3)$, $(M_{A_{12}}, X_4)$ and $(M_{A_{13}}, X_5)$ with parameters $\theta_1 = 2.264$, $\theta_2 = 1.664$ and $\theta_3 = 1.329$, respectively. Additionally, we used an asymmetric Gumbel copula for $(M_{A_{11}}, X_2)$ with parameters $\theta = 1.933$, $\alpha_1 = 0.900$ and $\alpha_2 = 1.000$. The $p$-values for the goodness of fit tests used for each copula were all larger than $0.17$.

c) The $p$-values from the comparison tests in the recursive validation also indicate that the copulas of the vectors from the original data seem to adequately match those of the sample generated with the Iman–Conover algorithm from the model.

Although the validation technique used on the aggregation-tree model is ad-hoc, its results illustrate the versatility of aggregation-tree models and their benefits over traditional copula models. With classical extreme-value copula models, we fail to obtain an adequate fit for this dataset. Since Gumbel copulas were mostly used for the aggregation-tree model, we attempted to fit a $5$-dimensional Gumbel copula model to the transformed

Fréchet data. Using the inverse $\tau$ methods, we obtain the parameter estimate $\hat{\theta} = 1.916$ for the $5$-dimensional copula. However, the $p$-value for a goodness-of-fit test for this $5$-dimensional model was quite small ($p = 0.0305$). This poor fit may be due in part to the inability of this model to account for certain aspects of the dependence structure of the data, such as the asymmetric dependence structure of the pair $(M_{A_{11}}, X_2)$. Indeed, as will be shown in Chapter 7, the Gumbel model enforces that any such pair has a bivariate Gumbel copula with the same parameter. By modelling the dependence structure between of each of these aggregated pairs directly, the aggregation-tree model allows us to capture the deeper intricacies of the dependence structure of the data.

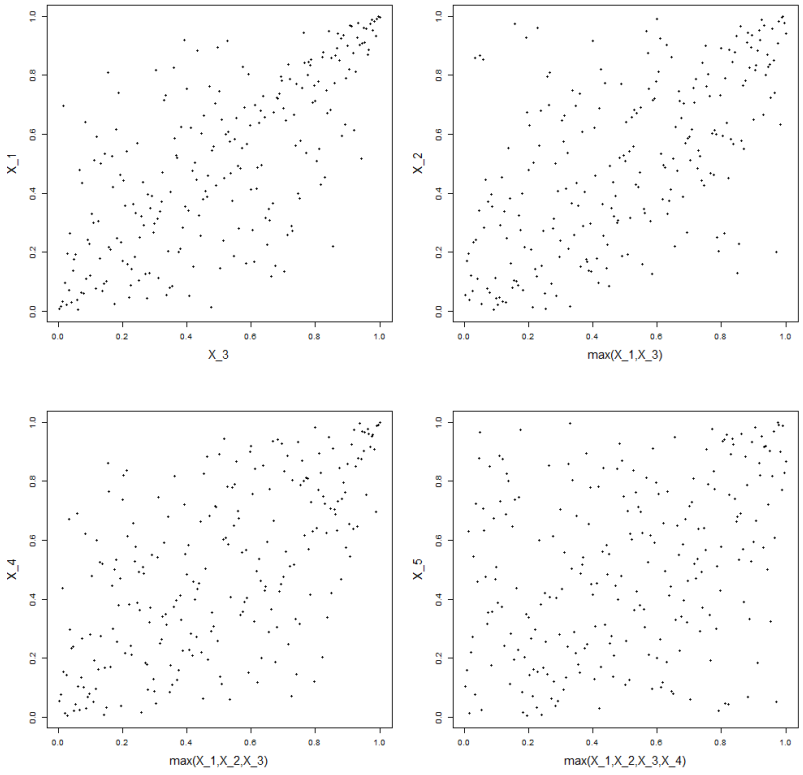


**Figure 6.5:** Rank plots of the sample generated using the Iman–Conover algorithm.

# Chapter 7

# Max-scale invariant aggregation-tree models

In this chapter, our purpose is to study the dependence structure of a random vector $(X_1, \ldots, X_d)$ derived from an aggregation-tree model in which only extreme-value copulas are used, and in which $a(x, y) = \max(x, y)$ serves as the aggregation function. At the outset, it is clear that it is possible to generate a multivariate extreme-value distribution using this approach. A trivial example would be an aggregation-tree model that uses bivariate extreme-value copulas to join distinct pairs of leaf nodes and independence copulas for all the other branching nodes.

However, one may suspect that other options exist, based on the data illustration in Chapter 6, where only extreme-value copulas were used in the construction of the aggregation-tree model. Indeed, when we applied a test for max-stability from Kojadinovic et al. (2011) for the copula of the sample generated from the model defined in Chapter 6, we obtained a very large $p$-value ($p = 0.9585$), which suggests that the overall copula may be extreme-value. In what follows, we present partial results relating to the conditions to ensure that the resulting joint distribution from an aggregation-tree model is multivariate extreme-value.

## 7.1 Max-scale invariance assumption

To simplify our investigation, we will limit ourselves herein to the case where the margins are unit Fréchet. Our motivation for doing so is that there is a particularly useful characterization for multivariate extreme-value vectors with unit Fréchet margins that we can exploit. This characterization, which is described below, is analogous to a characterization for min-stable random vectors from Weintraub (1991).

**Theorem 7.1.1.** *The joint distribution of a random vector $(X_1, \ldots, X_d)$ is extreme-value with Fréchet margins with shape parameter $\alpha = 1$ and location parameter $\beta = 0$ if and only if for any $a_1, \ldots, a_d \in [0, \infty)$ such that $a_j \neq 0$ for at least one $j \in \{1, \ldots, d\}$, the random variable $Y = \max(a_1 X_1, \ldots, a_d X_d)$ has a Fréchet distribution with shape parameter $\alpha = 1$ and location parameter $\beta = 0$. The random variable $Y$ is called a max-linear transformation of $(X_1, \ldots, X_d)$.*

*Proof.* First suppose that $\boldsymbol{X} = (X_1, \ldots, X_d)$ is extreme-value and suppose that for each $j \in \{1, \ldots, d\}$, $X_j$ is Fréchet with shape parameter $\alpha = 1$, location $\beta = 0$ and some scaling parameter $s_j \in (0, \infty)$. This means that the random vector $(X_1/s_1, \ldots, X_d/s_d)$ is then multivariate extreme-value with unit Fréchet margins. By characterization *(iii)* of Theorem 5.1.3, there exist non-negative functions $f_1, \ldots, f_d$ integrating to 1 such that, for all $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \mathbb{P}(X_1/s_1 \leq x_1/s_1, \ldots, X_d/s_d \leq x_d/s_d) = \exp\left\{ -\int_0^1 \bigvee_{i=1}^{d} \frac{s_i f_i(u)}{x_i} \, du \right\}.$$

Given that every subvector of $(X_1, \ldots, X_d)$ is also extreme-value, we need only consider the case where $a_1, \ldots, a_d \in (0, \infty)$. Let $Y = \max(a_1 X_1, \ldots, a_d X_d)$. For all $x \in (0, \infty)$, one then has

$$\begin{aligned} \mathbb{P}(Y \leq x) &= \mathbb{P}(a_1 X_1 \leq x, \ldots, a_d X_d \leq x) \\ &= \mathbb{P}\left( X_1 \leq \frac{x}{a_1}, \ldots, X_d \leq \frac{x}{a_d} \right) = \exp\left\{ -\frac{1}{x} \int_0^1 \bigvee_{i=1}^{d} s_i a_i f_i(u) du \right\}. \end{aligned}$$

Therefore, $Y$ is a Fréchet random variable with shape parameter $\alpha = 1$, location parameter $\beta = 0$, and scale parameter $s = \int_0^1 \bigvee_{i=1}^d s_i a_i f_i(u) du$.

Conversely, suppose that any max-linear transformation of $(X_1, \ldots, X_d)$ is a Fréchet random variable with shape $\alpha = 1$ and location $\beta = 0$. Fixing an arbitrary $j \in \{1, \ldots, d\}$ and setting $a_i = 0$ if $i \neq j$ and $a_j = 1$ then implies that $X_j$ is Fréchet with parameters $\alpha = 1$, $\beta = 0$, and some scaling $s_j \in (0, \infty)$.

By assumption, we know that, for all $x \in (0, \infty)$ and $(a_1, \ldots, a_d) \in [0, \infty)^d$ such that $a_j \neq 0$ for at least one $j \in \{1, \ldots, d\}$,

$$\mathbb{P}\left(\bigvee_{i=1}^d a_i X_i \leq x\right) = \exp\{-s(a_1, \ldots, a_d)/x\}$$

for some scale parameter $s(a_1, \ldots, a_d)$ depending on $a_1, \ldots, a_d$. For any $x_1, \ldots, x_d \in (0, \infty)$, we thus have in particular that

$$\begin{aligned}\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) &= \mathbb{P}\left(\frac{X_1}{x_1} \leq 1, \ldots, \frac{X_d}{x_d} \leq 1\right) \\ &= \mathbb{P}\left(\bigvee_{i=1}^d \frac{X_i}{x_i} \leq 1\right) = \exp\{-s(1/x_1, \ldots, 1/x_d)\}.\end{aligned}$$

Let $n \in \mathbb{N}$ be an integer. We can then write

$$\begin{aligned}\mathbb{P}^n(X_1 \leq nx_1, \ldots, X_d \leq nx_d) &= \mathbb{P}^n\left(\bigvee_{i=1}^d \frac{X_i}{x_i} \leq n\right) \\ &= \left[\exp\{-s(1/x_1, \ldots, 1/x_d)/n\}\right]^n \\ &= \exp\{-s(1/x_1, \ldots, 1/x_d)\} \\ &= \mathbb{P}\left(\bigvee_{i=1}^d \frac{X_i}{x_i} \leq 1\right) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d).\end{aligned}$$

Let $C$ be the unique copula of $\boldsymbol{X}$. Given that, for each $i \in \{1, \ldots, d\}$, $X_i$ has a scaled Fréchet distribution, one can write, $F_i(x) = e^{-s_i/x}$ for all $x \in (0, \infty)$. We have, for all

$x_1, \ldots, x_d \in (0, \infty)$,

$$\begin{aligned} C\{F_1(x_1), \ldots, F_d(x_d)\} &= \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) \\ &= \{\mathbb{P}(X_1 \leq nx_1, \ldots, X_d \leq nx_d)\}^n \\ &= C^n\big\{e^{-s_1/(nx_1)}, \ldots, e^{-s_d/(nx_d)}\big\} \\ &= C^n\big\{F_1^{1/n}(x_1), \ldots, F_d^{1/n}(x_d)\big\}. \end{aligned}$$

This implies that the copula of $\boldsymbol{X}$ is max-stable. Therefore, $\boldsymbol{X}$ is extreme-value with Fréchet margins. □

This result is useful for the articulation of additional assumptions concerning the aggregation-tree model which would ensure that the joint distribution is extreme-value, as we now discuss. To fit an aggregation-tree model to $\boldsymbol{X} = (X_1, \ldots, X_d)$ using the maximum as the aggregation function, we make assumptions about the copula of each

$$(M_{B_{1i}}, M_{B_{2i}}) = \left( \bigvee_{j \in B_{1i}} X_j, \bigvee_{j \in B_{2i}} X_j \right).$$

We can extend this assumption to a scaled version of $(X_1, \ldots, X_d)$ to ensure that the random vector $(X_1, \ldots, X_d)$ has an extreme-value distribution.

**Theorem 7.1.2.** *Consider an aggregation-tree model for a random vector for $(X_1, \ldots, X_d)$ with Fréchet margins with common shape parameter $\alpha = 1$ and location parameter $\beta = 0$. Suppose that the aggregation-tree model has the property that at each aggregation step $i$ in which the variables $M_{B_{1i}}$ and $M_{B_{2i}}$ are being combined, the copula of $(\bigvee_{j \in B_{1i}} a_j X_j, \bigvee_{j \in B_{2i}} a_j X_j)$ is extreme-value for any $a_1, \ldots, a_d \in [0, \infty)$ such that $a_j \neq 0$ for at least one $j \in B_{1i}$ and one $j \in B_{2i}$. Then the distribution of $(X_1, \ldots, X_d)$ is multivariate extreme-value.*

*Proof.* We proceed by induction on dimension $d \in \mathbb{N}$.

**Base case** ($d = 2$): The only possible tree structure for $d = 2$ variables has $B_{11} = \{1\}$ and $B_{21} = \{2\}$. By assumption, upon setting $a_1 = a_2 = 1$, the copula of $(X_1, X_2)$ is extreme-

value. Moreover, $X_1$ and $X_2$ have Fréchet distributions. Therefore, the distribution of $(X_1, X_2)$ is indeed multivariate extreme-value.

**Induction step**: Suppose that the statement holds for every tree structure with $k$ leaf nodes, whatever $k \in \{2, \ldots, d-1\}$. Thus, for every branching node $I \in \mathcal{B}(\mathcal{T})$, $(X_i : i \in I)$ has a multivariate extreme-value distribution. Let $B_1$ and $B_2$ be the children nodes of the root node. Then $B_1 \cup B_2 = \{1, \ldots, d\}$ and $\boldsymbol{X}_{B_1} = (X_j : j \in B_1)$ and $\boldsymbol{X}_{B_2} = (X_j : j \in B_2)$ are multivariate extreme-value random vectors with Fréchet margins with common shape $\alpha = 1$ and location $\beta = 0$ by the induction hypothesis. By Theorem 7.1.1, for any $a_1, \ldots, a_d \in [0, \infty)$ such that $a_j \neq 0$ for at least one $j \in B_1$ and one $j \in B_2$, $\bigvee_{j \in B_1} a_j X_j$ and $\bigvee_{j \in B_2} a_j X_j$ are both Fréchet distributed with shape parameter $\alpha = 1$, location parameter $\beta = 0$, and scaling parameters $s_1$ and $s_2$, respectively, both of which depend on $a_1, \ldots, a_d$.

Now observe that if $(Y, Z)$ is a bivariate extreme-value random vector with unit Fréchet margins and an extreme-value copula with Pickands dependence function $\mathcal{A}$, then for all $a, b \in (0, \infty)$, one has

$$\mathbb{P}\{\max(aY, bZ) \leq u\} = C(e^{-a/u}, e^{-b/u}) = \exp\left\{-\frac{a+b}{u}\,\mathcal{A}\left(\frac{a}{a+b}\right)\right\}. \tag{7.1}$$

Now consider $a_1, \ldots, a_d \in [0, \infty)$, where $a_j \neq 0$ for at least one $j$. If $a_j = 0$ for all $j \in B_1$, then $\bigvee_{j=1}^{d} a_j X_j = \bigvee_{j \in B_2} a_j X_j$. As explained above, we already know that $\bigvee_{j \in B_2} a_j X_j$ is Fréchet with shape $\alpha = 1$ and location $\beta = 0$. Similarly, if $a_j = 0$ for all $j \in B_2$, $\bigvee_{j=1}^{d} a_j X_j = \bigvee_{j \in B_1} a_j X_j$, where the latter is Fréchet with shape $\alpha = 1$ and $\beta = 0$. Therefore, it remains to consider the case in which $a_j \neq 0$ for at least one $j \in B_1$ and one $j \in B_2$. By assumption, the copula of

$$\left(\bigvee_{j \in B_1} a_j X_j, \bigvee_{j \in B_2} a_j X_j\right)$$

is extreme-value. Therefore, if $\mathcal{A}_a$ denotes its Pickands dependence function, we have from (7.1) that

$$\mathbb{P}\left(\bigvee_{j=1}^{d} a_j X_j \leq u\right) = \mathbb{P}\left(\bigvee_{j\in B_1} a_j X_j \leq u, \bigvee_{j\in B_2} a_j X_j \leq u\right)$$
$$= \exp\left\{-\frac{s_1+s_2}{u}\,\mathcal{A}_a\left(\frac{s_1}{s_1+s_2}\right)\right\}.$$

This implies that $\bigvee_{j=1}^{d} a_j X_j$ is Fréchet with shape $\alpha = 1$, location $\beta = 0$ and scaling $(s_1+s_2)\mathcal{A}_a\{s_1/(s_1+s_2)\}$. In conclusion, given that $\bigvee_{j=1}^{d} a_j X_j$ is Fréchet for all $a_1, \ldots, a_d \in [0,\infty)$ such that $a_j \neq 0$ for some $j \in \{1,\ldots,d\}$, $(X_1,\ldots,X_d)$ is multivariate extreme-value by Theorem 7.1.2. □

Observe that in order to ensure the property of the aggregation-tree model required in Theorem 7.1.2, it is necessary that all the copulas joining the nodes must be extreme-value. Below, we explore the use of a more restrictive assumption that also ensures that we obtain an extreme-value copula through the aggregation-tree model by Theorem 7.1.2.

**Max-scale invariance assumption**: Consider an aggregation-tree model based on the aggregation function $a(x,y) = \max(x,y)$, tree structure $\mathcal{T}$, and copulas $C_1,\ldots,C_{d-1}$ for a random vector $(X_1,\ldots,X_d)$. Let $M_{B_{1i}}$ and $M_{B_{2i}}$ be the variables aggregated at the $i$th aggregation step. It is assumed that, for all $i \in \{1,\ldots,d-1\}$ and for all $a_1,\ldots,a_d \in [0,\infty)$ such that $a_j \neq 0$ for at least one $j \in B_{1i}$ and at least one $j \in B_{2i}$, the copula that joins $(\bigvee_{j\in B_{1i}} a_j X_j, \bigvee_{j\in B_{2i}} a_j X_j)$ is $C_i$.

Although this max-scale invariance assumption may seem unrealistic, the next result shows that it is actually verified for multivariate extreme-value distributions with the Gumbel copula and Fréchet margins. However, it remains unclear whether such distributions are actually aggregation-tree models.

**Theorem 7.1.3.** *Let $\boldsymbol{X} = (X_1,\ldots,X_d)$ be a random vector with unit Fréchet margins whose underlying copula is Gumbel with parameter $\theta$. Then for any sets $A, B \subset \{1,\ldots,d\}$ with $A \cap B \neq$*

*$\emptyset$ and any $a_1, \ldots, a_d, b_1, \ldots, b_d \in [0, \infty)$ such that $a_j \neq 0$ for at least one $j \in A$ and $b_j \neq 0$ for at leat one $j \in B$, the copula of $(\bigvee_{j \in A} a_j X_j, \bigvee_{j \in B} b_j X_j)$ is bivariate Gumbel with parameter $\theta$.*

*Proof.* If the copula of $\boldsymbol{X}$ is a $d$-dimensional Gumbel with parameter $\theta$, then the copulas of $(X_j : j \in A)$ and $(X_j : j \in B)$ are both Gumbel with parameter $\theta$. By Theorem 7.1.1, we know that $\bigvee_{j \in A} a_j X_j$ and $\bigvee_{j \in B} b_j X_j$ are both Fréchet with shape $\alpha = 1$, location $\beta = 0$, and a scaling parameter dependent on the vectors $(a_1, \ldots, a_d)$ and $(b_1, \ldots, b_d)$, respectively. First, we find these scaling parameters. To this end, fix $x \in (0, \infty)$. Then

$$\begin{aligned}\mathbb{P}\left(\bigvee_{i \in A} a_i X_i \leq x\right) &= \mathbb{P}(a_i X_i \leq x : i \in A) \\ &= \exp\left[-\left\{\sum_{i \in A}\left(\frac{a_i}{x}\right)^\theta\right\}^{1/\theta}\right] = \exp\left\{-\frac{1}{x}\left(\sum_{i \in A} a_i^\theta\right)^{1/\theta}\right\}\end{aligned}$$

and similarly for $\bigvee_{j \in B} b_j X_j$. Therefore, the scaling parameters are

$$s_a = \left(\sum_{i \in A} a_i^\theta\right)^{1/\theta} \quad \text{and} \quad s_b = \left(\sum_{i \in B} b_i^\theta\right)^{1/\theta}$$

for $\bigvee_{j \in A} a_j X_j$ and $\bigvee_{j \in B} b_j X_j$, respectively.

Next, we examine the joint distribution of $(\bigvee_{j \in A} a_j X_j, \bigvee_{j \in B} b_j X_j)$. We have, for arbitrary $x, y \in (0, \infty)$,

$$\begin{aligned}\mathbb{P}\left(\bigvee_{i \in A} a_i X_i \leq x, \bigvee_{j \in B} b_j X_j \leq y\right) &= \mathbb{P}(\{a_i X_i \leq x : i \in A\} \cap \{b_i X_i \leq y : i \in B\}) \\ &= \exp\left[-\left\{\sum_{i \in A}\left(\frac{a_i}{x}\right)^\theta + \sum_{i \in B}\left(\frac{b_i}{y}\right)^\theta\right\}^{1/\theta}\right] \\ &= \exp\left[-\left\{\left(\frac{s_a}{x}\right)^\theta + \left(\frac{s_b}{y}\right)^\theta\right\}^{1/\theta}\right],\end{aligned}$$

which shows that the copula of $(\bigvee_{j \in A} a_j X_j, \bigvee_{j \in B} b_j X_j)$ is bivariate Gumbel with parameter $\theta$. This concludes the argument. □

The above result shows that to obtain a multivariate extreme-value distribution with a Gumbel copula with parameter $\theta$ from an aggregation-tree model, it is necessary to use only bivariate Gumbel copulas with the parameter $\theta$ at each aggregation step.

Our final result, stated below, explains why using only bivariate Gumbel copulas with the same parameter is sufficient to generate a multivariate extreme-value distribution with a Gumbel copula from an aggregation-tree model satisfying the max-scale invariance assumption.

**Theorem 7.1.4.** *Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be random vectors with Fréchet margins with shape parameter $\alpha = 1$ and location parameter $\beta = 0$. Let the copulas of $\boldsymbol{X}$ and $\boldsymbol{Y}$ be Gumbel with parameter $\theta$. Suppose that for any $a_1, \ldots, a_n, b_1, \ldots, b_m \in (0, \infty)$, the copula of $(\bigvee_{j=1}^n a_j X_j, \bigvee_{j=1}^m b_j Y_j)$ is bivariate Gumbel with parameter $\theta$. Then the copula of $(\boldsymbol{X}, \boldsymbol{Y}) = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is also Gumbel with parameter $\theta$.*

*Proof.* For any $\boldsymbol{x} = (x_1, \ldots, x_n) \in (0, \infty)^n$ and $\boldsymbol{y} = (y_1, \ldots, y_m) \in (0, \infty)^m$, consider $\bigvee_{j=1}^n X_j/x_j$ and $\bigvee_{j=1}^m Y_j/y_j$, which have Fréchet distributions with shape $\alpha = 1$ and location $\beta = 0$ by Theorem 7.1.1. From the proof of Theorem 7.1.3, the scaling parameters are $(\sum_{i=1}^n x_i^{-\theta})^{1/\theta}$ and $(\sum_{i=1}^m y_i^{-\theta})^{1/\theta}$, respectively.

We examine the joint distribution of $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$. For any $x_1, \ldots, x_n \in (0, \infty)$ and $y_1, \ldots, y_m \in (0, \infty)$, we have

$$\begin{aligned}\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n, Y_1 \le y_1, \ldots, Y_m \le y_m) &= \mathbb{P}\left(\bigvee_{j=1}^n \frac{X_j}{x_j} \le 1, \bigvee_{j=1}^m \frac{Y_j}{y_j} \le 1\right) \\ &= \exp\left[-\left\{\sum_{i=1}^n \left(\frac{1}{x_i}\right)^\theta + \sum_{i=1}^m \left(\frac{1}{y_i}\right)^\theta\right\}^{1/\theta}\right].\end{aligned}$$

This shows that the copula of $(\boldsymbol{X}, \boldsymbol{Y})$ is Gumbel with parameter $\theta$. □

Applying this result recursively at every aggregation step shows that an aggregation-tree model with aggregation function $a(x, y) = \max(x, y)$ which uses bivariate Gumbel

copulas with parameter $\theta$ and satisfies the max-scale invariance assumption generates a multivariate extreme-value distribution whose copula is Gumbel with parameter $\theta$.

The main obstacle that remains at present is the fact that it is unclear how to enforce the max-scale invariance assumption for aggregation-tree models. It also seems difficult to verify whether a multivariate extreme-value distribution can be expressed as an aggregation-tree model.

That said, observe that the multivariate Gumbel model requires us to use only the same copula for every pair of variables being aggregated. Under the max-scale invariance assumption, other types of copulas can perhaps be obtained when using different bivariate copulas across the aggregation process. However, it seems that this assumption may impose restrictions on the choice of copulas that can be used to model the aggregated pairs. For instance, consider a $4$-dimensional aggregation-tree model with tree structure $\mathcal{T}_2 = \{\{1,2\},3,4\}$, $\mathcal{T}_3 = \{\{1,2\},\{3,4\}\}$, $\mathcal{T}_4 = \{\{1,2,3,4\}\}$, and copulas $C_{\{1\}\{2\}}$, $C_{\{3\}\{4\}}$ and $C_{\{1,2\}\{3,4\}}$, which are Gumbel with parameters $\theta_1$, $\theta_2$, and $\theta_3$, respectively. If the max-scale invariance assumption holds, this model produces the $4$-dimensional copula defined, for all $u_1, \ldots, u_4 \in (0,1)$, by

$$C(u_1,u_2,u_3,u_4) = \exp[-\big\{(u_1^{\theta_1}+u_2^{\theta_1})^{\theta_3/\theta_1} + (u_3^{\theta_2}+u_4^{\theta_2})^{\theta_3/\theta_2}\big\}^{1/\theta_3}]. \tag{7.2}$$

The copula defined in (7.2) is known as a nested Gumbel copula. However, it is a valid copula only if $\theta_1 \leq \theta_3$ and $\theta_2 \leq \theta_3$; see Hofert & Pham (2012). Thus, it seems that the validity of the max-scale invariance assumption does somehow limit the choice of copulas used in the aggregation steps.

It also remains to be seen how to develop inference procedures which could be used to verify the validity of the max-scale invariance assumption. It would also be important to explore other possible copulas that could be obtained under the max-scale invariance assumption, and the limitations on the choice of copulas that the max-scale invariance assumption imposes.

# Chapter 8

# Conclusion

In this thesis, aggregation-tree models were seen to offer a simple and practical framework to build dependence structures for high-dimensional data.

We examined the necessary conditional independence assumption to ensure that the aggregation-tree model provides a unique and well-defined joint distribution, and we presented an alternative formulation which could be simpler to verify than those given by Arbenz et al. (2012) and Côté & Genest (2015). We also explored some of the applications and properties of aggregation-tree models for modeling extreme values.

Furthermore, we proposed a model selection technique adapted for extreme-value analysis using hierarchical clustering based on upper-tail dependence measures through the $F$-madogram. Under strong assumptions on the model construction, we found that aggregation-tree models based on extreme-value copulas and the maximum as the aggregating function can be used to generate multivariate extreme-value distributions.

As we saw, the max-scale invariance assumption stands out as a strong but simple condition which ensures that an aggregation-tree model generates a multivariate extreme-value distribution. This assumption is necessary to obtain certain multivariate extreme-value distributions such as those having a Gumbel copula and Fréchet margins. However, ways to enforce this assumption for aggregation-tree models remain to be found. More-

over, this assumption seems to place limitations on the choice of bivariate copulas that one can use. It would be interesting to articulate these limitations in future work.

# Bibliography


Arbenz, P., Hummel, C. & Mainik, G. (2012). Copula based hierarchical risk aggregation through sample reordering. *Insurance: Mathematics and Economics*, vol. 51, no. 1, pp. 122–133.

Ben Ghorbal, N., Genest, C. & Nešlehová, J. (2009). On the test of Ghoudi, Khoudraji, and Rivest for extreme-value dependence. *The Canadian Journal of Statistics*, vol. 37, no. 4, pp. 534–552.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. & Terpenning, I.J. (1990). STL: A seasonal-trend decomposition procedure based on LOESS. *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–33.

Cooley, D., Naveau, P. & Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in Probability and Statistics* (Bertail, P., Soulier, P. & Doukhan, P., eds). Lecture Notes in Statistics, vol. 187. Springer, pp. 373–390.

Côté, M.-P. & Genest, C. (2015). A copula-based risk aggregation model. *The Canadian Journal of Statistics*, vol. 43, no. 1, pp. 60–81.

Czado, C. (2019). *Analyzing Dependent Data With Vine Copulas: A Practical Guide With R*. Springer, New York.

Czado, C. & Nagler, T. (2021). Vine copula based modeling. *Annual Review of Statistics and Its Application*, vol. 9, pp. 453–477.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180–190.

Genest, C., Ghoudi, K. & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, vol. 82, no. 3, pp. 543–552.

Genest, C., Ghoudi, K. & Rivest, L.-P. (1998). Comment on "Understanding relationships using copulas" by E.W. Frees and E.A. Valdez. *North American Actuarial Journal*, vol. 2, no. 1, pp. 143–149.

Genest, C., Kojadinovic, I., Nešlehová, J.G. & Yan, J. (2011). A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, vol. 17, no. 1, pp. 253–275.

Genest, C., Mesfioui, M. & Nešlehová, J.G. (2019). On the asymptotic covariance of the multivariate empirical copula process. *Dependence Modeling*, vol. 7, no. 1, pp. 279–291.

Genest, C. & Nešlehová, J. (2007). A primer on copulas for count data. *The ASTIN Bulletin*, vol. 37, no. 2, pp. 475–515.

Genest, C. & Nešlehová, J. (2012). Copulas and copula models. *Encyclopedia of Environmetrics* (El-Shaarawi, A.H. & Piegorsch, W.W.). Wiley, Chichester, vol. 2, pp. 541–553.

Genest, C. & Nešlehová, J. (2013). Assessing and modeling asymmetry in bivariate continuous data. In *Copulae in Mathematical and Quantitative Finance* (Jaworski, P., Durante, F. & Härdle, W.K., eds). Springer, Berlin, pp. 91–114.

Genest, C., Nešlehová, J. & Ben Ghorbal, N. (2010). Spearman's footrule and Gini's gamma: A review with complements. *Journal of Nonparametric Statistics*, vol. 22, no. 8, pp. 937–954.

Genest, C., Nešlehová, J. & Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 4, pp. 811–834.

Genest, C. & Rémillard, B. (2004). Tests of independence and randomness based on the empirical copula process. *TEST*, vol. 13, no. 2, pp. 335–369.

Genest, C., Rémillard, B. & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 199–214.

Genest, C. & Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, vol. 37, no. 5B, pp. 2990–3022.

Genest, C. & Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *Journal of Multivariate Analysis*, vol. 101, no. 8, pp. 1837–1845.

Ghoudi, K., Khoudraji, A. & Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *The Canadian Journal of Statistics*, vol. 26, no. 1, pp. 187–197.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, vol. 44, no. 3, pp. 423–453.

Grønneberg, S. & Hjort, N.L. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, vol. 41, no. 2, pp. 436–459.

Hofert, M., Kojadinovic, I., Maechler, M. & Yan, J. (2023). `copula`: Multivariate dependence with copulas. R package version 1.1-2.

Hofert, M. & Pham, D. (2012). Densities of nested Archimedean copulas. *Journal of Multivariate Analysis*. vol. 118, pp. 37–52.

Iman, R.L. & Conover, W.J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics, Simulation and Computation*, vol. 11, no. 3, pp. 311–334.

Jaworski, P., Durante, F., Härdle, W.K. & Rychlik, T. (2010). *Copula Theory and Its Applications*, Proceedings of the Workshop Held in Warsaw, 25-26 September 2009. Springer, Berlin, Heidelberg.

Joe, H. (2014). *Dependence Modeling With Copulas*. CRC Press, Boca Raton, FL.

Kallenberg, O. (2022). *Foundations of Modern Probability*, 3rd ed. Springer, New York.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivariées*. PhD dissertation, Université Laval, Québec, QC, Canada.

Kojadinovic, I., Segers, J. & Yan, J. (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *The Canadian Journal of Statistics*, vol. 39, no. 4, pp. 703–720.

Kojadinovic, I. & Yan, J. (2012). A non-parametric test of exchangeability for extreme -value and left-tail decreasing bivariate copulas. *Scandinavian Journal of Statistics*, vol. 30, no. 3, pp. 480–496.

Kurowicka, D. & Joe, H. (2010). *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing, Hackensack, NJ.

McNeil, A.J., Frey, R. & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ.

McNeil, A.J. & Nešlehová, J. (2009). Multivariate Archimedean copulas, $d$-monotone functions and $\ell_1$-norm symmetric distributions. *The Annals of Statistics*, vol. 37, no. 5B, pp. 3059–3097.

Murphy, O.A. (2018). *Extremal Modeling of Dependent and Non-Stationary Series and Joint Modeling of Extreme Rainfall at Multiple Sites*. Doctoral dissertation, McGill University, Montréal, QC, Canada.

Naveau, P., Guillou, A., Cooley, D. & Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, vol. 96, no. 1, pp. 1–17.

Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York.

Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, vol. 49, pp. 859–878 and pp. 894–902.

Rémillard B. (2017). Goodness-of-fit tests for copulas of multivariate time series. *Econometrics*, vol. 5, no. 1, pp. 1–23.

Rémillard, B. & Scaillet, O. (2009). Testing for equality between two copulas. *Journal of Multivariate Analysis*, vol. 100, no. 3, pp. 377–386.

Rémillard, B. & Plante, J. (2012). `TwoCop`: Nonparametric test of equality between two copulas. R package version 1.0.

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, vol. 18, no. 3, pp. 764–782.

Sklar, A. (1959). Fonctions de répartition à $n$ dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, vol. 8, pp. 229–231.

Weintraub, K.S. (1991). Sample and ergodic properties of some min-stable processes. *The Annals of Probability*, vol. 19, no. 2, pp. 706–723.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press, Cambridge.