Inference and Sampling for Archimax Copulas

Anonymous Author(s)
Affiliation
Address
email

Abstract

Understanding multivariate dependencies in both the bulk and the tails of a distribution is an important problem for many applications, such as ensuring algorithms are robust to observations that are infrequent but have devastating effects. Archimax copulas are a family of distributions endowed with a precise representation that allows simultaneous modeling of the bulk and the tails of a distribution. Rather than separating the two as is typically done in practice, incorporating additional information from the bulk may improve inference of the tails, where observations are limited. Building on the stochastic representation of Archimax copulas, we develop a non-parametric inference method and sampling algorithm. Our proposed methods, to the best of our knowledge, are the first that allow for highly flexible and scalable inference and sampling algorithms, enabling the increased use of Archimax copulas in practical settings. We experimentally compare to state-of-the-art density modeling techniques, and the results suggest that the proposed method effectively extrapolates to tails while scaling to higher dimensional data. Our findings suggest that the proposed algorithms can be used in a variety of applications where understanding the interplay between the bulk and the tails of a distribution is necessary, such as healthcare and safety.

1 Introduction

Modeling dependencies between random variables is a central task in statistics, and understanding the dependence between covariates throughout the distribution is important in characterizing a distribution outside of its areas of highest density. For example, in machine learning contexts, a major topic of interest lies in enforcing dependencies between covariates, such as spatial dependence in convolutional neural networks or temporal dependence in recurrent neural networks. Copulas are functions that model the joint dependence of random variables, and they have been successfully employed in a variety of practical modeling settings due to their ease of use and intuitive properties.

Moreover, since copulas are used to represent cumulative distribution functions (CDFs), they have been particularly useful in situations where *tail events* are important – events that have high impact but low probability. For example, in computer vision applications, local dependence modeled by convolutions may be sufficient for the bulk of the data, but for data in the tails of the distribution non-local dependencies may be present and need to be modeled. Current successful applications of copulas have largely been limited to a few basic parametric families and low dimensional settings, preventing their widespread adoption in settings where the dependencies are complicated or the dimension is large.

Archimax copulas are a class of copulas that merges a tractable form with sufficient expressiveness. Archimax copulas effectively balance the representation of data within the bulk of the distribution while extrapolating to the tails by combining the tractability of *Archimedean* copulas with the tail properties of *extreme value copulas*.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

Notably, they remove the simplified symmetry assumption among covariates that is present in Archimedean copulas and the max-stable property, a result of which is lower tail independence, of extreme value copulas. These are major assumptions often not held in practice. The use of Archimax copulas has resulted in better fit to data in applications such as healthcare [25] and hydrology [3, 11].

However, existing computational methods do not allow feasible inference and sampling for Archimax copulas, preventing their widespread use. Therein lies the motivation behind this work. We construct efficient and flexible inference and sampling methods using deep learning techniques and discuss how they compare to traditional means of density estimation that use existing copula methods and deep generative models. In addition, we provide numerical studies where the proposed Archimax techniques extrapolate to the tails better than existing methods.

1.1 Related work

Modeling distributions is a major task in machine learning, with techniques such as generative adversarial networks (GANs) [12], normalizing flows (NFs) [30], and variational autoencoders (VAEs) [21] being the major developments for representing complex distributions.

However, these techniques are largely used for modeling the bulk of a distribution and may not extrapolate well to out-of-distribution samples or samples within the tails, as discussed in [23]. Some methods have been proposed to represent only the tails, for example in [1, 6]. However, these methods are used only for sampling, and do not provide a way of quantifying the dependence of the observation. Moreover, they disregard the information in the bulk of the distribution and only focus on the tails.

Copulas are an important technique for representing distribution functions since they allow easy separation of the marginals and the joint dependence structure of a distribution. Copulas have been applied in machine learning wherein techniques from machine learning have been used in conjunction with traditional copula theory to model more general classes of densities with examples of such work found in [24, 28, 35], please see Appendix A for more background on copulas and application of copulas in machine learning. However, these have generally focused on simplified assumptions such as a symmetric dependence between covariates, or a hierarchy of bivariate dependencies. Moreover, these do not extrapolate to tail distributions and do not readily appear to generalize to high dimensions.

With regards to Archimax copulas, several theoretical works have been proposed analyzing the distributions [9, 10, 27]. In Chatelain et al. [11], the authors proposed a method for inferring a stable tail dependence function (stdf) when given an Archimedean generator. However the method assumes knowledge of the Archimedean generator or infers a one-parameter Archimedean generator from pairwise Kendall's taus. Past applications of Archimax copulas were of low dimensions, such as dimensions 2, 3 and 3 in the studies of river flow rates [3], rainfall [11] and nutrient intake [25].

Our contributions We propose methods for filling the gaps in the existing literature by

1. Developing methods for inferring both the Archimedean generator and stdf;
2. Developing methods for sampling from Archimax copulas;
3. Providing flexible representations for both the radial and spectral components.

2 Background

Copulas are given by separating the marginal distributions from the joint dependence of a random variable. Specifically, consider

$$F(\mathbf{x}) = C(F_1(x_1), \cdots, F_d(x_d)), \tag{1}$$

where F is a d-variate cumulative distribution function (CDF) of the random variable $\mathbf{X} = (X_1, \cdots, X_d) \in \mathbb{R}^d$, F_j is the jth univariate margin, and C is the copula describing the dependence between the uniform random variables $\mathbf{U} = (U_1, \cdots, U_d) = (F_1(X_1), \cdots, F_d(X_d)) \in [0, 1]^d$. Moreover, if the marginals F_j are continuous, then the copula C is unique [34].[1]

[1] notation: random variables in uppercase, observations in lowercase, scalars are not bold, vectors are bold.

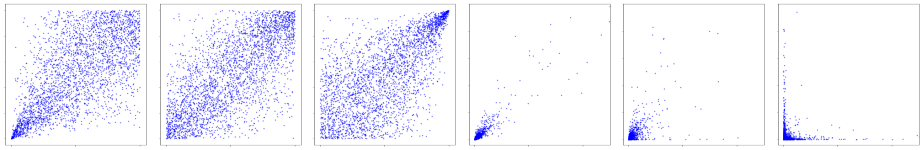

Figure 1: Examples of different radial envelopes and asymptotic dependencies.

As discussed in the introduction, Archimax copulas describe a generalization of Archimedean and extreme value copulas. They are defined as:

Definition 1 (Archimax copula) *An Archimax copula is given by*

$$C(\mathbf{u}) = \varphi(\ell(\varphi^{-1}(u_1), \cdots, \varphi^{-1}(u_d))), \tag{2}$$

where $\varphi : [0, \infty) \to [0, 1]$ *is an* Archimedean generator *and* $\ell : [0, \infty)^d \to [0, \infty)$ *is a* stable tail dependence function (stdf) *[8, 10, 27].*

From the definition, the Archimax copula is completely characterized by these two functions and the objective during inference lies in estimating both the *stdf* and the *Archimedean generator*. Both functions have specific properties that must be fulfilled in order to obtain a valid Archimax copula.

The *stdf* is defined as:

Definition 2 (stable tail dependence function (stdf)) *A* d*-variate stdf* $\ell : [0, \infty)^d \to [0, \infty)$ *is given by:*

$$\ell(\mathbf{x}) = d \int_{\Delta_{d-1}} \max_{j \in \{1, \cdots, d\}} \{x_j w_j\} \, \mathrm{d}F_{\mathbf{w}}(\mathbf{w}) = d\, \mathbb{E}_{\mathbf{W}} \left[\max_{j \in \{1, \cdots, d\}} \{x_j W_j\} \right], \tag{3}$$

with a spectral random variable $\mathbf{W} \in \Delta_{d-1}$ *satisfying moment constraints* $\mathbb{E}_{\mathbf{W}}[W_j] = 1/d$ *for* $j \in \{1, \cdots, d\}$ *[13, 20, 33].*

Intuitively, the *stdf* dictates the asymptotic dependence between covariates. Notably, the definition completely relies on the distribution of $\mathbf{W}$, and the *stdf* is homogeneous of order one, i.e. $\ell(cx_1, \cdots, cx_d) = c\, \ell(x_1, \cdots, x_d)$ for $c > 0$.

On the other hand, the *Archimedean generator* is defined by:

Definition 3 (Archimedean generator) *An Archimedean generator* $\varphi : [0, \infty) \to [0, 1]$ *is the Williamson* d*-transform of a random variable* $R > 0$*:*

$$\varphi(x) = \mathcal{W}_R(x) = \int_x^\infty \left(1 - \frac{x}{r}\right)^{d-1} dF_R(r) = \mathbb{E}_R \left[\left(1 - \frac{x}{R}\right)_+^{d-1} \right], \tag{4}$$

where $(y)_+ := \max(0, y)$ *for* $y \in \mathbb{R}$ *[26, 37].*

The Archimedean generator has a one-to-one correspondence with the distribution of R [26] and dictates the shape of the radial envelope applied across all covariates.

With these definitions in mind, we leverage Charpentier et al. [10, Theorem 3.3] for inference and sampling, which established that any random variable

$$\mathbf{X} \stackrel{d}{=} R \times (S_1, \cdots, S_d), \tag{5}$$

follows an Archimax copula, where R and $\mathbf{S}$ are independent and known as the *radial* and *simplex* components, with supports $R > 0$, $\mathbf{S} \in [0, 1]^d$, $\ell(\mathbf{S}) \leq 1$. Moreover, every Archimax copula given in (2) has a decomposition given by (5). The *radial* component R is the same R in Definition 3, and the *simplex* component $\mathbf{S}$ has a one-to-one correspondence with the *spectral* component $\mathbf{W}$ from Definition 2. Therefore, representing the R and $\mathbf{W}$ in Definitions 3 and 2 respectively provides complete understanding of the distribution.

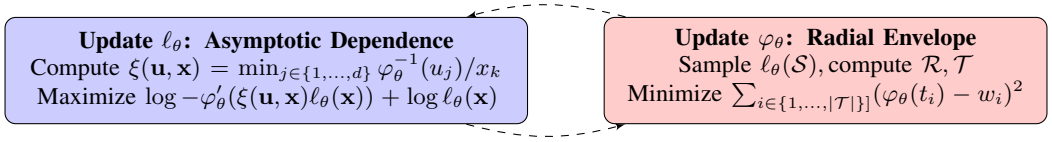

Figure 2: Flow chart describing relationship between stdf and Archimedean generator estimation.

3 Method

As established in Section 2, the *stdf* and *Archimedean generator* are expectations of functions of the *spectral* component $\mathbf{W}$ and *radial* component R respectively. We first model these expectations where $\mathbf{W}$ and R are discrete random variables with finite support [15, 17]. We then let them be outputs of generative networks in the limit of infinite support. We begin by describing the inference algorithm for the stdf followed by the inference algorithm for the Archimedean generator. Through sampling the stdf we define the relationship between $\mathbf{S}$ and $\mathbf{W}$ which we make use of in the inference for the Archimedean generator. We finally show how the representations we use can be easily adapted for sampling from Archimax copulas. A flow chart describing the relationship between stdf and Archimedean generator parameter estimation is in Figure 2. The source code with detailed documentation will be made available online.

3.1 Stable tail dependence function inference and sampling

For this section, we suppose that the Archimedean generator $\varphi(\cdot)$ is known and the goal is to infer or sample from the stdf.

3.1.1 Inference for stable tail dependence function

We consider a stdf following Definition 2, where ℓ is specified by the spectral decomposition with spectral component $\mathbf{W}$ [13, 20, 33]. To summarize the inference for the stdf: We first establish a representation of the stdf through the empirical expectation of samples $\mathbf{w}$ from a generative network $G_{\mathbf{W}}$. We then define the likelihood of transformed data observations and optimize the parameters of $G_{\mathbf{W}}$ to maximize this likelihood.

Following Chatelain et al. [11], when given the Archimedean generator φ we can define the transformation $\xi(\mathbf{u},\mathbf{x})$ for an observation $\mathbf{u}$ and a pseudo-observation $\mathbf{x} = (x_1, \cdots, x_d) \in \Delta_{d-1}$:

$$\xi(\mathbf{u},\mathbf{x}) = \min\{\varphi^{-1}(u_1)/x_1, \cdots, \varphi^{-1}(u_d)/x_d\}, \tag{6}$$

a transformation used for estimating extreme value copulas [8, 31].

In the case of Archimax copulas, and by using the homogeneity property of the stdf (3), the CDF of the random variable $\xi(\mathbf{U},\mathbf{x})$ is expressed as:

$$P(\xi(\mathbf{U},\mathbf{x}) \leq x) = 1 - P(\xi(\mathbf{U},\mathbf{x}) > x) = 1 - C(\varphi(x\mathbf{x})) = 1 - \varphi(x\ell(\mathbf{x})). \tag{7}$$

The full derivation is given in Appendix A.

Following Hasan et al. [18], we differentiate the CDF of $\xi(\mathbf{U},\mathbf{x})$ with respect to x and obtain the log-likelihood of the transformed observation as:

$$\log \mathcal{L}(x;\varphi,\ell) = \log\left(-\varphi'(x\,\ell(\mathbf{x}))\right) + \log(\ell(\mathbf{x})). \tag{8}$$

Recalling Definition 2, let ℓ_θ be parameterized by a generative network $G_{\mathbf{W}}$ such that:

$$\ell_\theta(\mathbf{x}) = \frac{d}{l}\sum_{k=1}^{l}\left[\max_{j\in\{1,\cdots,d\}} x_j w_{kj}\right], \tag{9}$$

where $\mathbf{w}_k = (w_{k1}, \cdots, w_{kd})$ for $k = \{1, \cdots, l\}$ are l samples from $G_{\mathbf{W}}$ with dimension d and output activation given by the Softmax$(\cdot)$ function to respect the support $\mathbf{W} \in \Delta_{d-1}$. The moment constraints $\mathbb{E}_{\mathbf{W}}[W_j] = 1/d$ for $j = \{1, \cdots, d\}$ are scale conveniences and not necessities [15]. To approximate the moment constraints, we penalize the residual $\sum_{j=1}^{d}(\sum_{k=1}^{l} w_{kj}/l - 1/d)^2$.

We may then train the generative network $G_{\mathbf{W}}$ such that ℓ_θ approximates ℓ by minimizing the negative log-likelihood of transformed observations $\xi_(\mathbf{u}, \mathbf{x})$, with $\mathbf{x}$ uniformly sampled on the unit simplex Δ_{d-1}. The full technique is presented in Algorithm 1 in Appendix A.

The inverse φ^{-1} in the computation of $\xi(\mathbf{u}, \mathbf{x})$ in (6) can be computed numerically by Newton-Raphson. The derivative φ' in the computation of $\log \mathcal{L}(x; \varphi, \ell)$ in (8) can be calculated explicitly from Definition 3 as:

$$\varphi'(x) = \mathbb{E}_R \left[(d-1) \left(-\frac{1}{R} \right) \left(1 - \frac{x}{R} \right)_+^{d-2} \right]. \tag{10}$$

Putting these computations together results in the likelihood of the stdf given the generator.

3.1.2 Sampling the simplex component

We now consider sampling the simplex component $\mathbf{S}$. We noted in the introduction that there is a one-to-one relationship between $\mathbf{S}$ and $\mathbf{W}$ which we explicitly note here. Using this relationship and the result of Charpentier et al. [10], the stdf may also be written in terms of the simplex component through the following definition:

Definition 4 (survival distribution of simplex component) *The survival distribution function (SDF) of the simplex component* $\mathbf{S}$ *is:*

$$P(\mathbf{S} > \mathbf{s}) = \max(0, 1 - \ell(s_1, \cdots, s_d))^{d-1}, \tag{11}$$

which relates to the so-called generalized Pareto copulas defined as:

Definition 5 (generalized Pareto copula) *A generalized Pareto copula is defined as the copula of a generalized Pareto distribution with support on* $(-\infty, 0]^d$ *and CDF specified by* ℓ *as:*

$$P(X_1 < x_1, \cdots, X_d < x_d) = \max(0, 1 - \ell(-x_1, \cdots, x_d)). \tag{12}$$

Moreover, it has stochastic representation as:

$$\mathbf{X} \stackrel{d}{=} -\left(\frac{U}{W_1}, \cdots, \frac{U}{W_d} \right), \tag{13}$$

where U *is uniformly distribution on* $[0, 1]$*, independent of spectral component* $\mathbf{W} = (W_1, \cdots, W_d)$ *from the spectral decomposition of* ℓ *in Definition 2 [7, 14].*

Therefore, we can frame the sampling of the simplex component through the sampling from a generalized Pareto copula, such that, given samples $(x_{11}, \cdots, x_{1d}), \cdots, (x_{(d-1)1}, \cdots, x_{(d-1)d})$ of size $d-1$ from a generalized Pareto distribution with CDF in (12), we may obtain a sample of $\mathbf{s}$, where coordinates s_j for $j = \{1, \cdots, d\}$ are computed as [10]:

$$s_j = -\max(x_{1j}, \cdots, x_{(d-1)j}). \tag{14}$$

The coordinate-wise maxima is taken over $d-1$ samples to get the $d-1$ exponent in the SDF of $\mathbf{S}$, corresponding to the $d-1$ exponent in the expression of the Archimedean generator φ.

Due to our convenient representation of ℓ as an expectation of a function of $\mathbf{W}$ from generative network $G_{\mathbf{W}}$, we are able to sample a generalized Pareto distribution from its stochastic representation. The full technique is presented in Algorithm 2 in Appendix A.

3.2 Archimedean generator inference and sampling

We now assume that the stdf is known and the goal is to infer the Archimedean generator. We consider a general representation of the Archimedean generator following Definition 3, where φ is a $d-$monotone function specified by the Williamson $d-$transform of the radial component R [10, 26].

To provide an outline for the overall approach, we first consider the so-called Kendall distribution, which describes an integral transform of the copula. We then use the fact that the empirical Kendall distribution converges to the true Kendall distribution as the empirical copula converges to the true copula, providing a means of estimation. The Kendall distribution is formally defined as:

Definition 6 (Kendall distribution) *Let C be a copula, and let the CDF of the random variable $\mathbf{U} \in [0,1]^d$ be the copula C. Define the random variable $W := C(\mathbf{U})$. The Kendall distribution of C is the multivariate probability integral transform given by the CDF of W:*

$$K(w) = P(C(\mathbf{U}) \leq w), \quad w \in [0,1]. \tag{15}$$

In the case of Archimax copulas, using the stochastic representation (5) and the homogeneity property of the stdf (3), the Kendall distribution is expressed as:

$$K(w) = P(\varphi(R\,\ell(S_1, \cdots, S_d)) \leq w). \tag{16}$$

The full derivation is provided in Appendix A. The empirical Kendall distribution K_n for n given observations $(u_{11}, \cdots, u_{1d}), \cdots, (u_{n1}, \cdots, u_{nd})$ is defined as:

$$K_n(w) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{w_i \leq w\}, \tag{17}$$

where

$$w_i = \frac{1}{n+1} \sum_{k=1}^{n} \mathbb{1}\{u_{k1} < u_{i1}, \cdots, u_{kd} < u_{id}\}. \tag{18}$$

In the case of symmetric dependence in the non-parametric inference of Archimedean copulas [17], $\mathbf{S}$ is uniformly distributed on the simplex Δ_{d-1} and $\ell(S_1, \cdots, S_d) = S_1 + \cdots + S_d \equiv 1$. Then, R is discrete with the cardinality of the support the same as W and K_n is the Kendall distribution of a unique Archimedean copula [17].

In the non-parametric inference of Archimax copulas, we define the random variables:

$$Z := \ell(S_1, \cdots, S_d) \quad \text{and} \quad T := RZ \tag{19}$$

where T is a discrete random variable with support the same size as W and K_n is the Kendall distribution of an Archimax copula. Note that R and Z are independent as R and $\mathbf{S}$ are independent.

Using K_n, we reconstruct the support of R, given ℓ and the support of $\mathbf{S}$, which in turn provides the support of Z. With the final objective of letting R and $\mathbf{S}$ be independent and identically distributed (iid) outputs of generative networks, we first (linearly) interpolate K_n to be equispaced, such that each w_i has the same probability $k_i = 1/(n_r n_z)$, where n_r and n_z are the chosen sizes of supports for R and Z. We describe the reconstruction procedure for the general case with non-iid random variables including cases where $r_j z_l = r_{j'} z_{l'}$ in Appendix A. The sizes of supports n_r and n_z are chosen empirically with examples given in Appendix B.

Suppose the supports of the distributions of R, Z, W, T are finite and respectively denoted by: $\mathcal{W} = \{w_1, w_2, \cdots, w_{n_r n_z}\}$, $\mathcal{R} = \{r_1, r_2, \cdots, r_{n_r}\}$, $\mathcal{Z} = \{z_1, z_2, \cdots, z_{n_z}\}$, $\mathcal{T} = \{r_j z_l : r_j \in \mathcal{R}, z_l \in \mathcal{Z}\} = \{t_1, t_2, \cdots, t_{n_r n_z}\}$, where the $n_r n_z$ elements of $\mathcal{W}$ are sorted in decreasing (non-increasing) order, and the $n_r n_z$ elements of $\mathcal{T}$ are sorted in increasing (non-decreasing) order. This reverse ordering is due to φ being a decreasing function.

We minimize the mean sum of square residuals motivated by the uniform convergence of the empirical process $\sqrt{n}(K_n - K), n \to \infty$ established by Barbe et al. [4]:

$$\frac{1}{n_r n_z} \sum_{i=1}^{n_r n_z} (w_i - \varphi(t_i))^2 \tag{20}$$

where, following the definition of φ in (4),

$$\varphi(t_i) = \mathcal{W}_{\mathcal{R}}(t_i) = \frac{1}{n_r} \sum_{j=1}^{n_r} \left(1 - \frac{t_i}{r_j}\right)_+^{d-1}. \tag{21}$$

The finite support assumption of R and Z is not necessary, and we consider a modification where the supports $\mathcal{R}, \mathcal{Z}$ are specified by samples from generative networks. The main objective is to learn the parameters of generative network G_R given the empirical Kendall distribution K_n and samples of Z. Since scaling the support $\mathcal{R}$ by a constant $c > 0$ does not change the copula, we add a regularization term for $\mathbb{E}_R[R] = 1$. The algorithm can be understood as an alternating minimization algorithm, where the map between $\mathcal{R}$ and $\mathcal{Z}$ to $\mathcal{W}$ via $\mathcal{T}$, and the support $\mathcal{R}$ are updated in an alternating fashion, please see Algorithm 3 in Appendix A for details.

The learned G_R also provides a source of samples for the radial component R, which we will use to generate full samples from the Archimax copula.

3.3 Inference and sampling for Archimax copulas

We finally summarize the combination of inference and sampling for both the Archimedean and simplex component to obtain the full algorithm for inference and sampling for Archimax copulas. This culminates into an iterative technique that successively updates each component.

3.3.1 Inference for Archimax copulas

We initialize with $\varphi(x) = \exp(-x)$ and infer ℓ to learn $G_{\mathbf{W}}$, following Section 3.1. This special combination of φ and ℓ corresponds to extreme-value dependence with the max-stable property.To aid inference for this initialization step, we pre-process our data to have extreme-value dependence. We do so by computing the *block maxima*, a technique from extreme value copulas where we group observations and take the coordinate-wise maximas within each group. Specifically, given observations $(x_{11}, \cdots, x_{1d}), \cdots, (x_{n1}, \cdots, x_{nd})$, the block-maximas $(m_{11}, \cdots, m_{1d}), \cdots, (m_{k1}, \cdots, m_{kd})$ for k blocks of size n/k are computed as:

$$m_{ij} = \max(x_{(n/k)(i-1)+1\,j}, \cdots, x_{(n/k)(i)\,j}), \tag{22}$$

for $i = \{1, \cdots, k\}$ and $j = \{1, \cdots, d\}$. To determine the block size, with larger block sizes n/k better approximating extreme-value dependence and more blocks k for more observations, we test the block maximas for extreme-value dependence via the max-stable property using the test by Kojadinovic et al. [22], and select the first k where the null hypothesis of extreme-value dependence is not rejected, starting from $k = n$.

Given an estimate of ℓ and a learned $G_{\mathbf{W}}$, we are able to generate many samples of $\mathbf{S}$, thereby providing a way to compute $Z = \ell(\mathbf{S})$. We then infer φ, learning G_R, following Section 3.2. At this point, we repeat the estimation for $G_{\mathbf{W}}$ with the updated φ to improve our estimate for ℓ. The algorithms may be iterated as needed with suggested convergence criteria such as Cramér–von Mises (CvM) distance between successively estimated copulas.

3.3.2 Sampling for Archimax copulas

Given samples of the radial and simplex components, the stochastic representation of Archimax copulas (5) gives a straightforward method for sampling Archimax copulas [10]. Specifically, we simply sample $\mathbf{S}$ and R and multiply to obtain a sample from the Archimax copula.

4 Experiments

To understand the modeling and sampling capabilities of the proposed algorithms, we conduct a number of empirical studies. We compare the proposed method to a number of existing copula-based and deep learning-based methods for density estimation. Our experiments relate to the main focus of the proposed method where we are interested in understanding the dependencies between the variables in both the bulk and the tail. In that sense, we conduct a number of experiments where we wish to extrapolate to the tail. Further experimental details may be found in Appendix B.

The metric we use to compare the methods is based on the Cramér-von Mises (CvM) statistic, which computes the L^2 distance between the empirical copula and the estimated copula. This statistic is commonly used to determine differences between distributions in goodness-of-fit tests [32]. We use a version where we compare the empirical copula of true samples to the empirical copula of generated samples. Specifically, it can be written as: $CvM = \int (C_{*,n}(\mathbf{u}) - C_{\theta,n}(\mathbf{u}))^2 \mathrm{d}\mathbf{u}$, where $C_{*,n}$ is the empirical copula of true samples and $C_{\theta,n}$ is the empirical copula of generated samples [32].

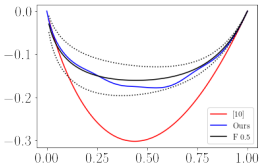

Figure 3: Estimate of $\lambda(w)$: blue (our proposed method), red [11], black (ground truth).

Inference for Archimedean generator Our first set of experiments involve inference for the Archimedean generator following the proposed method in Section 3.2. We consider data with dimension $d = 10$ and sample size of $n = 1000$ given the true stdf ℓ. To the best of our knowledge, our proposed method is the first method for non-parametric inference of flexible Archimedean generators in Archimax copulas. As such, there are no baselines for comparison. Instead, we compute the results in terms

of the map $\lambda(w) = \varphi^{-1}(w)/(\varphi^{-1}(w))', w \in (0,1)$ due to its scale invariant property and known asymptotic variance, which was described as a useful metric for how well φ is fit in [16, 17]. We evaluated our proposed method on the Clayton, Frank, Joe and Gumbel generators, representing different radial envelopes, for Kendall's tau of $\tau = \{0.2, 0.5\}$, representing different associations. The stdf comes from the family of negative scaled extremal Dirichlet (NSD), a flexible class of stdf [5]. All estimates were within the asymptotic variance of $\lambda(w), w \in (0,1)$. The next best method is comparing to a Clayton generator estimated by pairwise Kendall's tau [11]. We give the results in terms of MSE to λ in Table 1, and plot estimates of $\lambda(w)$ in Figure 3. Additional experiment results, including small sample performance $n = 200$, and choices of support sizes n_r and n_z, are in Appendix B.

Inference for stable tail dependence function and sampling for simplex component We now consider the opposite case where we wish to estimate the stdf ℓ given the true Archimedean generator φ. We compute the integrated relative absolute error (IRAE) between the estimated ℓ_θ and the true ℓ. The IRAE is given by $IRAE(\ell, \ell_\theta) = \frac{1}{|\Delta_{d-1}|} \int_{\Delta_{d-1}} |\ell(\mathbf{x}) - \ell_\theta(\mathbf{x})|/\ell(\mathbf{x})\, \mathrm{d}\mathbf{x}$ [11]. For the same settings as above, results given true φ are in Table 1.

Table 1: Inference of φ given true ℓ and inference of ℓ given true φ

	C 0.2	C 0.5	F 0.2	F 0.5	J 0.2	J 0.5	G 0.2	G 0.5
MSE $\times 10^{-3}$ [11]	**0.01**	**0.04**	0.9	9	1	9	0.7	8
MSE $\times 10^{-3}$ (OURS)	0.2	0.2	**0.1**	**0.1**	**0.3**	**0.1**	**0.2**	**0.1**
IRAE ± 0.01 [11]	0.05	0.11	0.04	0.04	0.05	1.00	0.06	0.15
IRAE ± 0.01 (OURS)	0.06	0.12	0.04	0.05	0.06	**0.07**	0.08	0.15

Modeling nutrient intake The USDA studied the nutrient intake of women [36]. One particular task is understanding the dependencies between the intake of different nutrients. We can model this using an Archimax copula to understand the dependencies between the nutrients in the bulk and the tail. To assess this, we fit and compare models from the literature on representing distributions and compute the CvM goodness-of-fit statistic for each model. Specifically, we break our comparison into two different types of models: copula based models and deep network based models. For the copula models, we use the same marginal distributions for all methods including the proposed method but consider various forms for the copula parameterization. For methods marked with *, we use the φ described in Section 3.2 and for methods marked with † we use the sampling method described in Section 3.1.2. For the deep generative models, we use standard methods based on the Wasserstein GAN [2], masked autoregressive flow (MAF) [29], and variational autoencoders (VAE) [21]. The results are presented in Table 2. The proposed method (Gen-AX) with the Archimax has the lowest CvM distance among all the competing methods, suggesting that the proposed method is recovering the true dependency structure. The state of the art Clayton Archimax (C-AX) did not perform well possibly due to the difficulty in scaling the single parameter Clayton generator to higher dimension. The Archimedean copula (AC *) possibly benefited from the use of our proposed φ. We additionally provide examples of samples versus the ground truth in Appendix B for the different methods as well as an explanation of the different abbreviations.

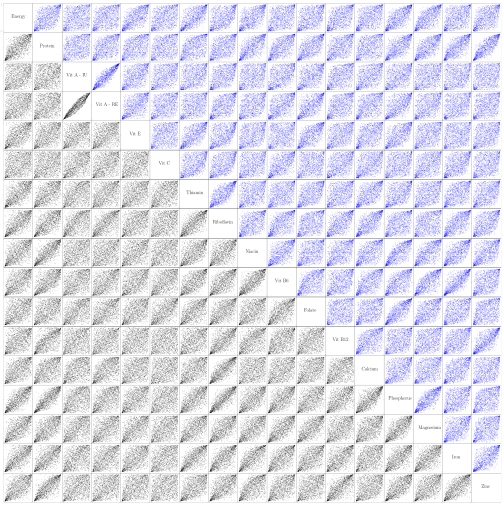

Figure 4: Nutrient intake samples in blue are the proposed Archimax method, samples in black are the real samples.

Extrapolating to extreme rainfall We also consider another realistic dataset which models the rainfall in French Britanny as studied by Chatelain et al. [11]. In this experiment, we are interested in

Table 2: Goodness-of-fit to nutrient intake data and 100d NSD copula

	COPULAS								DEEP NETS			OURS
	GC	RV	CV	DV	AC *	HAC	EV †	C-AX †	GAN	MAF	VAE	Gen-AX *†
Nutrient CvM $\times 10^{-3}$	0.081	0.3	0.2	0.6	0.030	0.2	0.4	0.6	0.033	0.036	0.053	**0.026**
C-NSD CvM $\times 10^{-5}$	-	-	-	-	-	-	-	-	16	**3**	5	**3**

testing how well the proposed method can extrapolate to the extremes from non-extreme data. This dataset consists of 3 locations (Belle-Ile, Groix, and Lorient) where the rainfall was recorded over the time period. For this experiment, we first train the models on the full dataset. We then generate many samples from the trained model, compute the block maxima, then estimate the extremal dependence from the block maxima using the CFG estimator [8]. The results are presented in Table 3 and Figure 5 where we compare the proposed method to deep generative models. We did not compare to copula based models since classical copulas are generally Gumbel or independent in the extremes and not suitable for this application.

Table 3: Goodness-of-fit to dependence in the extremes for rainfall data.

IRAE $\pm$ 0.01	GAN	MAF	VAE	C-AX	OURS
C-NSD	0.52	0.42	0.16	0.12	**0.03**
F-NSD	0.10	0.15	**0.04**	**0.04**	**0.04**
J-NSD	**0.03**	0.48	**0.03**	0.08	**0.03**
G-NSD	0.07	0.38	0.08	0.16	**0.04**

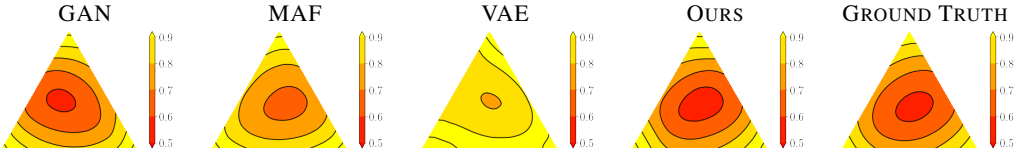

Figure 5: Extrapolating to dependence in the extremes for rainfall data. Plots of $\ell(\mathbf{w}), \mathbf{w} \in \Delta_{3-1}$.

High dimensional modeling We finally consider an experiment where we sample data from a 100-dimensional Clayton-NSD Archimax copula. Scaling to high dimensions is an important property of the proposed method since many existing copula methods fail to scale beyond lower dimensions. As such, we only compare with deep generative models since the existing copula models resulted in numerical errors during optimization. We report the CvM statistic in the second line of Table 2. We additionally provide examples of the samples in Appendix B.

5 Conclusion

We developed highly flexible and scalable inference and sampling algorithms, facilitating the use of Archimax copulas in practical settings. We experimentally compare to state-of-the-art density modeling techniques, and the results suggest that the proposed method effectively extrapolates to tails while scaling to higher dimensional data. The methods are especially useful in scenarios requiring extrapolations to the tails while also incorporating data from the bulk.

Limitations and future work A single Archimedean generator φ to describe the radial envelope across all coordinates may not be sufficiently expressive for certain datasets. For these cases, hierarchical Archimax copulas may be more appropriate and a direction of future work [19].

Potential negative societal impacts Model misspecification may lead to misspecification of risks, leading to potentially catastrophic outcomes in areas such as healthcare, safety and finance. Risks may be mitigated by confirming a reasonable fit between the observations generated by the model and the data.

References

[1] Michaël Allouche, Stéphane Girard, and Emmanuel Gobet. Tail-gan: Simulation of extreme events with relu neural networks. 2021.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[3] Tomáš Bacigál, Vladimír Jágr, and Radko Mesiar. Non-exchangeable random variables, archimax copulas and their fitting to real data. *Kybernetika*, 47(4):519–531, 2011.

[4] Philippe Barbe, Christian Genest, Kilani Ghoudi, and Bruno Rémillard. On kendall's process. *Journal of Multivariate Analysis*, 58(2):197–229, 1996.

[5] Léo R. Belzile and Johanna G. Nešlehová. Extremal attractors of liouville copulas. *Journal of Multivariate Analysis*, 160:68–92, 2017. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2017.05.008. URL `https://www.sciencedirect.com/science/article/pii/S0047259X17300453`.

[6] Younes Boulaguiem, Jakob Zscheischler, Edoardo Vignotto, Karin van der Wiel, and Sebastian Engelke. Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks. *Environmental Data Science*, 1:e5, 2022. doi: 10.1017/eds.2022.4.

[7] T. A. Buishand, L. de Haan, and C. Zhou. On spatial extremes: With application to a rainfall problem. *The Annals of Applied Statistics*, 2(2):624–642, 2008. ISSN 19326157. URL `http://www.jstor.org/stable/30244220`.

[8] P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 09 1997.

[9] Philippe Capéraà, Anne-Laure Fougères, and Christian Genest. Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*, 72(1):30–49, 2000.

[10] A. Charpentier, A.-L. Fougères, C. Genest, and J.G. Nešlehová. Multivariate archimax copulas. *Journal of Multivariate Analysis*, 126:118–136, 2014.

[11] Simon Chatelain, Anne-Laure Fougères, and Johanna G. Nešlehová. Inference for Archimax copulas. *The Annals of Statistics*, 48(2):1025 – 1051, 2020.

[12] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.

[13] L. De Haan. A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4):1194 – 1204, 1984.

[14] Michael Falk, Simone A. Padoan, and Florian Wisheckel. Generalized pareto copulas: A key to multivariate extremes. *Journal of Multivariate Analysis*, 174:104538, 2019. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2019.104538. URL `https://www.sciencedirect.com/science/article/pii/S0047259X19300296`.

[15] Anne-Laure Fougères, Cécile Mercadier, and John Nolan. Dense classes of multivariate extreme value distributions. *Journal of Multivariate Analysis*, 116:109–129, 2013. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2012.11.015. URL `https://www.sciencedirect.com/science/article/pii/S0047259X12002746`.

[16] Christian Genest and Louis-Paul Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043, 1993.

[17] Christian Genest, Johanna Nešlehová, and Johanna Ziegel. Inference in multivariate archimedean copula models. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 20(2):223, 2011.

[18] Ali Hasan, Khalil Elkhalil, Yuting Ng, Joao M. Pereira, Sina Farsiu, Jose H. Blanchet, and Vahid Tarokh. Modeling extremes with d-max-decreasing neural networks. In *38th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2022.

[19] Marius Hofert, Raphaël Huser, and Avinash Prasad. Hierarchical archimax copulas. *Journal of Multivariate Analysis*, 167:195–211, 2018.

[20] Xu Huang. *Statistics of bivariate extreme values*. PhD, Tinbergen Institute, 1992.

[21] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[22] Ivan Kojadinovic, Johan Segers, and Jun Yan. Large-sample tests of extreme-value dependence for multivariate copulas. LIDAM Reprints ISBA 2011025, Université catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA), 2011. URL `https://EconPapers.repec.org/RePEc:aiz:louvar:2011025`.

[23] Charline Le Lan and Laurent Dinh. Perfect density models cannot guarantee anomaly detection. *Entropy*, 23(12):1690, 2021.

[24] Chun Kai Ling, Fei Fang, and J. Zico Kolter. Deep archimedean copulas. In *Advances in Neural Information Processing Systems*, volume 33, pages 1535–1545. Curran Associates, Inc., 2020.

[25] Alexander J McNeil and Johanna Nešlehová. From archimedean to liouville copulas. *Journal of Multivariate Analysis*, 101(8):1772–1790, 2010.

[26] Alexander J. McNeil and Johanna Nešlehová. Multivariate Archimedean copulas, d-monotone functions and $l1$-norm symmetric distributions. *The Annals of Statistics*, 37(5B):3059 – 3097, 2009.

[27] Radko Mesiar and Vladimír Jágr. d-dimensional dependence functions and archimax copulas. *Fuzzy Sets and Systems*, 228:78–87, 2013.

[28] Yuting Ng, Ali Hasan, Khalil Elkhalil, and Vahid Tarokh. Generative archimedean copulas. In *37th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021.

[29] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.

[30] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

[31] James Pickands. Multivariate extreme value distribution. In *43rd Session of the International Statistical Institution*, volume 2, pages 859–878,894–902, Buenos Aires, 1981.

[32] Bruno Remillard and Olivier Scaillet. Testing for equality between two copulas. 8:229–231, june 2007. doi: http://dx.doi.org/10.2139/ssrn.1014550. URL `https://ssrn.com/abstract=1014550`.

[33] Paul Ressel. Homogeneous distributions—and a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis*, 117:246–256, 2013.

[34] Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. 8:229–231, 1959.

[35] Natasa Tagasovska, Damien Ackerer, and Thibault Vatter. Copulas as high-dimensional generative models: Vine copula autoencoders. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper/2019/file/15e122e839dfdaa7ce969536f94aecf6-Paper.pdf`.

[36] USDA. Dataset: CSFII 1985, continuing survey of food intakes by individuals, women 19-50 years of age and their children 1-5 years of age, 6 waves, 1985. Dataset, U.S. Department of Agriculture, Agriculture Research Service, 1985. `https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/8586/csfii85_6waves_doc.pdf`, `https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/csfii-1985-1986/`.

[37] R. E. Williamson. Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal*, 23(2):189 – 207, 1956.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] See Appendix A.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Appendix B, and the code in the supplemental material is to be made available online.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Appendix B.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See Appendix B.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Appendix B.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] Nutrient intake data [36], Copulas `https://github.com/sdv-dev/Copulas`, HAC toolbox `https://github.com/gorecki/HACopula`, PyTorch, SciPy, NumPy, Python.
   (b) Did you mention the license of the assets? [No] Most information presented on the USDA Web site is considered public domain information. Public domain information may be freely distributed or copied, but use of appropriate byline/photo/image credits is requested. Attribution may be cited as follows: "U.S. Department of Agriculture."
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] The code in the supplemental material is to be made available online.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] Data was collected only after consent was given.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] Personally identifiable information has been removed.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [No] The instructions given to participants is available at [36].
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [No] The description of potential participant risks is available at [36].

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No] The description of participant compensation is available at [36].