

FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION  
FOR HIGHER PROFESSIONAL EDUCATION  
NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS  
*Faculty: St. Petersburg School of Economics and Management*

Vladimir Yashin

MASTER THESIS

**RECURSIVE NON-PARAMETRIC ESTIMATION  
OF THE COPULA DENSITY**

Field of study: 01.04.02 Applied Mathematics and Informatics

Degree program: Big Data Analysis in Business, Economy, and Society

Reviewer

---

---

---

Supervisor

Assistant Professor, Ph. D.  
Faculty of Computer Science  
Geoffrey Decrouez

Saint Petersburg  
2018

## Abstract

The contribution of this paper is two-fold. First of all, we introduced an algorithm that estimates the joint density function using the smoothed non-parametric recursive estimator. Secondly, for the first time, we present a smoothed non-parametric recursive estimator for a multivariate copula density, a joint distribution that can capture complex dependence structures among its variates. Also, we present the numerical results that show an apparent convergence of these estimators.

**Key words:** stochastic approximation algorithms, on-line estimation, copula density.

## 1 Introduction

Two parts explain the nature of a random vector: (a) the marginal behaviour of each component; and (b) the dependence structure that governs the association between these components. Both can be achieved using copulas (Sklar, 1959). Specifically, let  $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$  be a random vector with joint cumulative distribution function  $F$  and its continuous marginals  $F_1, \dots, F_d$ , given  $d \geq 2$ . The quantile of order  $u_j$  is defined as  $q^{u_j} := F_j^{-1}(u_j)$ , for  $u_j \in (0, 1)$ ,  $j = 1 \dots d$ . According to Sklar's theorem, the joint distribution of  $\mathbf{X}$  can be represented by means of a unique function  $C : (0, 1)^d \rightarrow (0, 1)$  such that

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) = F(q^{u_1}, \dots, q^{u_d}). \quad (1)$$

The function  $C$  is referred to as the copula. The proof of the existence of  $C$  was not been provided in Sklar, (1959) but has been provided later on in Moore and Spruill, (1975) and Deheuvels, (1978).

Copulas find several applications in many fields. First of all, use of copulas lies in finance and insurance risk management. The methodology that relies on the notion of the copulas was prevalent in finance in the early to mid 00's. Copulas are well suited for the field of financial markets as they are inherently volatile and erratic.

Some earlier applications involve modelling the default correlation between financial instruments. To measure the correlation, it is crucial to estimate the joint distribution for the survival times (time until default) of the instruments. The notion of copulas is quite useful in this case as they input marginal distributions that are generally can be derived from credit curves which are constructed using the information that is available on the market, and return the joint distribution of survival times (Li, 1999). Other authors claim that the model must be dynamic, which requires measuring default risks of other obligors on a daily basis (Schönbucher and Schubert, 2001).

Another group of authors aimed to find evidence of the dependence of credit risk in credit default swaps on stock performance. In particular, they were looking for tail dependence between markets. Since the marginal distributions are asymmetric and tail correlation is not linear, the copulas are used as they can capture non-linear relationship. (Andersen and Sidenius, 2004; Dupuis et al., 2009; Kresta and Tichý, 2012; Silva et al., 2014). For a more comprehensive review of how copulas are applied in finance a reader is referred to Cherubini et al., (2004), Patton, (2009), and Patton, (2012).

Besides, copulas are used in the fields of hydrology and climatology. Most hydrological events are multivariate. For example, climate events like storms and floods are usually described using several dependent random variables like volume, duration, peak etc. For more information of this application the reader is referred to (Chu, 2011; Dempster et al., 2007; Favre et al., 2004; Friedman and Huang, 2010; Genest et al., 2007; Piantadosi, Howlett, and Borwein, 2012; Piantadosi, Howlett, Borwein, and Henstridge, 2012; Renard and Lang, 2007; Salvadori et al., 2007).

Copulas also can be applied to sampling task including generation of stationary time-series. The main advantage of copulas in this regard is the fact that they are capable of accounting not only for correlations but also for the complete dependence structure, unlike other methods. See also (Erhardt and Czado, 2010, 2012; Hyrš and Schwarz, 2015; Matthias and Jan-frederik, 2017; Nelsen, 2007; Strelen, 2009; Strelen and Nassaj, 2007).

In this work, we introduce for the first time an algorithm of the estimation of a multivariate copula density function using joint distribution and marginal distribution functions. The joint and marginals, as well as the multivariate copula density, are estimated using the smoothing non-parametric recursive estimators.

This paper is divided into seven sections. Section 2 presents the review of the traditional batch estimation of the copula. In the third section, the recursive estimate of the quantile can be found as well as a historical preview of this topic. Section 4, we introduce the recursive estimation of the copula density. Then, we discuss how to select a proper bandwidth for the presented algorithms. Section 6 presents the numerical results for a set of copulas with different parameter as well as the different choice of bandwidths.

## 2 Traditional Ways of Copula Estimation

In this section, we present a comprehensive review of the previously developed traditional batch-based methods of the copula estimation.

Suppose that  $\mathbf{X}_1, \dots, \mathbf{X}_N$  is a sequence of i.i.d. random variables such that  $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,d}) \in \mathbb{R}^d$ . The first non-parametric approach to estimate  $C$  was introduced in Ruschendorf, (1976), using pseudo-observations.

Let  $C(u_1, \dots, u_d)$  be a copula and  $F$  is a cumulative distribution function, and  $F_1, \dots, F_d$  are its marginals

$$\begin{aligned} C(u_1, \dots, u_d) &= F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \\ &\approx \frac{1}{n} \sum_{i=1}^N \mathbf{1}(X_{i,1} \leq F_1^{-1}(u_1), \dots, X_{i,d} \leq F_d^{-1}(u_d)) \\ &= \frac{1}{n} \sum_{i=1}^N \mathbf{1}(F_1(X_{i,1}) \leq u_1, \dots, F_d(X_{i,d}) \leq u_d), \end{aligned} \tag{2}$$

where for each of the marginals  $F_j(X_{i,j}) \forall j \in 1 \dots d$

$$\begin{aligned} F_j(X_{i,j}) &= \mathbb{P}(\mathbf{X}_j \leq X_{i,j}) \\ &= \mathbb{P}((X_{i,1}, \dots, X_{i,d}) \leq X_{i,j}) \\ &\approx \frac{1}{n} \sum_{k=1}^N \mathbf{1}(X_{k,j} \leq X_{i,j}) \\ &=: X_{i,j}^*. \end{aligned}$$

The variable  $X_{i,j}^*$  represent the normalized rank of the  $j$ -th element of the  $i$ -th observation. Then, the copula estimator is

$$\hat{C}_n(\mathbf{u}) := \frac{1}{n} \sum_{i=1}^N \mathbf{1}(\mathbf{X}_i^* \leq \mathbf{u}),$$

where  $\mathbf{X}_i^* = (X_{i,1}^*, \dots, X_{i,d}^*)$  and  $\mathbf{u} = (u_1, \dots, u_d)$ . Alternatively, a copula may be estimated using empirical quantiles (Deheuvels, 1979). The empirical quintiles are usually defined as follows

$$Q_n^{u_j} := \inf_{\vartheta} \left\{ \frac{1}{n} \sum_{i=1}^N \mathbf{1}(X_{i,j} \leq \vartheta) \geq u_j \right\}.$$

Then, given  $F_j^{-1}(u_j) = Q_n^{u_j}$  we may obtain from (2) the estimator of copula which is simply

$$\tilde{C}_n(\mathbf{u}) := \frac{1}{n} \sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \leq \mathbf{Q}_n^{\mathbf{u}}),$$

where  $\mathbf{Q}_n^{\mathbf{u}} := (Q_n^{u_1}, \dots, Q_n^{u_d})$ .

Both  $\hat{C}_n$  and  $\tilde{C}_n$  need to be completely recalculated once a new observation is obtained. This makes it essentially useless in practice since data flow can have a quite large  $N$ . In this study, we purpose an on-line version of copula estimator that requires only  $\mathcal{O}(N)$  operations, i. e. making updates in  $\mathcal{O}(1)$  operations. This can be achieved using a recursive estimation of the quantiles.

### 3 Recursive Estimation of the Quantile

This section presents the perquisites of stochastic approximation methods as well as the application of this method specifically to the estimation of the quantile.

In general, the task is to find the root of a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ . The first idea is to use the Newton's procedure that involve calculation of  $f$  at some point and a first-order derivative

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}.$$

However,  $f$  is not necessarily differentiable. Thus, we may use a slightly less efficient algorithm which though addresses this issue. Generally it can be defined as follows

$$x_n = x_{n-1} - \alpha f(x_{n-1}), \quad (3)$$

where  $\alpha$  is a positive number that defines a step size. Then, let us assume that we do not know the analytical expression for  $f$  and only have noisy observations  $f(x_n) + \varepsilon_n$  where  $\mathbb{E}[\varepsilon_n] = 0$  for each  $n \in 1 \dots N$ . So,  $f(x_n)$  can be replaced by averaging many noisy observations such that

$$\begin{aligned} f(x_n) &\approx \frac{1}{m} \sum_{i=1}^m (f(x_i) + u_i) \\ &\equiv \frac{1}{m} \sum_{i=1}^m F(x_i, u_i). \end{aligned}$$

Then, (3) can be rewritten in the following way

$$x_n = x_{n-1} - \frac{\alpha}{m} \sum_{i=1}^m F(x_i, u_i).$$

Note that this algorithm is obviously inefficient as it is required to run it  $m$  times for every  $n$ . To solve this problem, Robbins and Monro, (1951) suggested to use the following

$$x_n = x_{n-1} - \alpha_n F(x_n, u_n),$$

where  $\alpha_n$  is the sequence of positive numbers that converge to 0 such that  $\sum_n \alpha_n = \infty$  and  $\sum_n \alpha_n^2 < \infty$ . Using the Robbins-Monro algorithm we can derive a recursive estimator of a quantile.

Let us take  $F(Q_n^{u_j}, X_{n,j}) = \mathbf{1}(X_{n,j} \leq Q_n^{u_j}) - u_j$  for  $u_j \in (0, 1)$ , also consider a continuous function  $G$  such that  $G(Q_n^{u_j})$  is well defined. Then,

$$\begin{aligned} f(x) &= \mathbb{E}[F(Q_n^{u_j}, X_{n,j})] \\ &= \mathbb{P}(X_{n,j} \leq Q_n^{u_j}) - u_j \\ &= G(Q_n^{u_j}) - u_j. \end{aligned}$$

The root of this function is exactly  $u_j$  quantile of  $G$  and, thus, can be written recursively as

$$Q_n^{u_j} = Q_{n-1}^{u_j} + \alpha_n \left( u_j - \mathbf{1}(X_{n,j} \leq Q_{n-1}^{u_j}) \right).$$

This estimator can be further applied for estimation of probability density function which is required in the recursive quantile estimator itself as well as in the copula density estimator.

The on-line estimation of a probability density were initially introduced in late 60's in Wolverton and Wagner, (1969) and studied in 70's by, for example, Yamato, (1971), Davies, (1973), Wegman and Davies, (1979). Later in late 00's similar results were presented from the

view of stochastic approximation in Mokkadem et al., (2009); the application of the presented estimators was introduced in Slaoui, (2014).

There two ways how to estimate a probability density function: parametric and non-parametric. The parametric approach relies on the idea that a random set of random variables  $X_1, \dots, X_N$  belongs a parametric family of distributions and these parameters can be estimated, for example, using maximum a posteriori hypothesis. Otherwise, the non-parametric approach has no assumptions regarding a distribution.

A non-parametric approach of the estimation of the quantile  $Q_n^{u_j}$  involves an indicator function, so that that the updates do not depend on how high the difference is. It does not matter how much the current estimation of the quantile is higher or lower than the current data point, the size of the increment will be the same. It is natural to introduce a kernel to address this issue (Amiri and Thiam, 2014). Though, for the simplicity, an indicator is used in this study. We may rewrite and define an algorithm for estimation of empirical quantiles using on-line updates as follows

$$f_{j,n}^{u_j} = \left(1 - \frac{1}{n}\right) f_{j,n-1}^{u_j} + \frac{1}{n} K_{h_{j,n}}^{u_j} (Q_{n-1}^{u_j} - X_{n,j}) \quad (4)$$

$$a_{j,n}^{u_j} = \max \left[ \mu_{j,n}, \min [f_{j,n}^{u_j}, \nu \ln(n+1)] \right] \quad (5)$$

$$Q_n^{u_j} = Q_{n-1}^{u_j} + \frac{1}{na_{j,n}^{u_j}} \left[ u_j - \mathbf{1}(X_{n,j} \leq Q_{n-1}^{u_j}) \right], \quad (6)$$

where  $f_{j,n}^{u_j}$  is the recursive estimator for the marginal on the  $j$ -th ( $j = 1 \dots d$ ) component of the joint density  $f$ . In addition,  $K_h(x) := h^{-1}K(x/h)$  where  $K$  is a symmetric, univariate kernel evaluated at given points of grid, and  $h$  is the bandwidth for the kernel  $K$ . Note that,  $K = \int_{\mathbb{R}} K(x)dx = 1$ ,  $\mu_{j,n}$  and  $\nu$  are some positive constants.

**Theorem 1** (Point-wise almost sure convergence). *If the following assumptions are held*

1.  $K$  is a Lipchitz continuous twice differentiable kernel such that  $\int_{\mathbb{R}} K(z) dz = 1$ ,  $\int_{\mathbb{R}} zK(z) dz = 0$ , and  $\eta = \int_{\mathbb{R}} z^2 K(z) dz \in (0, \infty)$ ;
2.  $f_j^{u_j}$  is bounded and can be continuously differentiated twice at any given  $q^{u_j}$ ;
3.  $h_n \rightarrow 0$  but  $nh_n \rightarrow \infty$  as  $n \rightarrow \infty$ .

then for any  $Q_n^{u_j}$  and  $f_{j,n}^{u_j}$  obtained from the algorithm which is presented in (4), (5), and (6) implies the almost sure convergence

- a.  $|Q_n^{u_j} - q^{u_j}| \rightarrow 0$ ;
- b.  $|a_{j,n}^{u_j} - f_{j,n}^{u_j}| \rightarrow 0$ .

For the proof of a more general version of the theorem a reader is refereed to Amiri and Thiam, (2014) and Camirand and Decrouez, (2018).

We now present the asymptotic normality of the estimator of the marginal density  $f_{j,n}^{u_j}$  with bandwidths  $h_{j,n}$  at the set (but not the grid) of points  $u_j$ .

**Theorem 2** (Asymptotic normality of the marginal density estimator). *Assuming that each of the following assumptions is satisfied*

1. *There exists  $R_1 \geq 0$ ,  $R_2 > 0$ , and  $\lambda \in (0, 1)$  such that  $nh_{j,n}^5 \rightarrow R_1$  and  $n^{1-\lambda}h_{j,n} \rightarrow R_2$ ;*
2.  *$K$  is a Lipchitz continuous twice differentiable kernel such that  $\int_{\mathbb{R}} K(z) dz = 1$ ,  $\int_{\mathbb{R}} zK(z) dz = 0$ , and  $\eta = \int_{\mathbb{R}} z^2 K(z) dz \in (0, \infty)$ ;*
3. *A sequence of bandwidths  $h_{j,n} \rightarrow 0$  and  $nh_{j,n} \rightarrow \infty$  as  $n \rightarrow \infty$ ,*

*then the sequence*

$$(nh_{j,n})^{1/2} \left[ f_{j,n}^{u_j} - f_j^{u_j} - \frac{R_1^{1/2} \cdot (f_j^{u_j})'' \eta}{2 - \lambda} \right]$$

*is asymptotically normally distributed with mean 0 and variance  $f_j^{u_j}(2 - \lambda)^{-1}\tilde{\eta}$ , given  $\tilde{\eta} := \int_{\mathbb{R}} K^2(z) dz$*

The proof of Theorem 2 is provided in the mentioned papers. In the next section, we present the recursive smooth non-parametric estimation of the copula density as well as a theorem that generalizes Theorem 2 for the multivariate case.

## 4 Recursive Estimation of the Copula Density

Consider the definition of a copula  $C$  given in (1). Then, the copula density is simply

$$c(u_1, \dots, u_d) := \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1, \dots, \partial u_d} = \frac{\partial^d F(q^{u_1}, \dots, q^{u_d})}{\partial u_1, \dots, \partial u_d} = \frac{f(q^{u_1}, \dots, q^{u_d})}{\prod_{j=1}^d f_j(q^{u_j})} = \frac{f^{\mathbf{u}}}{f_1^{u_1} \dots f_d^{u_d}},$$

where  $f^{\mathbf{u}}$  denotes the multivariate density of  $\mathbf{X}$  and  $f_j^{u_j}$  is a marginal of its the  $j$ -th component.

In order to calculate  $c(\cdot)$  it is required to estimate  $f^{\mathbf{u}}$  and  $f_j^{u_j}$ ,  $j = 1 \dots d$ . A recursive non-parametric estimator of  $f_j^{u_j}$  is provided in (4). It remains to find a similar one for  $f^{\mathbf{u}}$ . To achieve this, we follow the approach of Mokkadem et al., (2009) that used a stochastic approximation of a multivariate density that is based on a kernel density estimator.

So, we need to estimate  $f^{\mathbf{u}}$  given a set of quantiles  $q^{\mathbf{u}} = (q^{u_1}, \dots, q^{u_d})$

$$f_n^{\mathbf{u}} = \left(1 - \frac{1}{n}\right) f_{n-1}^{\mathbf{u}} + \frac{1}{n} W_{b_n}^{\mathbf{u}}(Q_{n-1}^{\mathbf{u}} - \mathbf{X}_n) \quad (7)$$

where  $W_{b_n}^{\mathbf{u}}(\mathbf{X}) := \prod_{j=1}^d K_{b_n}^{u_j}(\mathbf{X}_j)$ . In addition,  $Q_{n-1}^{\mathbf{u}}$  are quantiles estimated at the given grid of points  $\mathbf{u}$ , and  $\mathbf{X}_n$  is the current  $d$ -dimensional data point.

First, we need to state the asymptotic normality of the estimator  $f_n^{\mathbf{u}}$  at the grid of points  $\mathbf{u} \in (0, 1)^d$ .

**Theorem 3** (Asymptotic normality of the joint density estimator). *Consider a set of assumptions*

1.  $b_n \rightarrow 0$  but  $nb_n \rightarrow \infty$  as  $n \rightarrow \infty$ ;
2.  $K$  is a symmetric, Lipchitz continuous, and compactly supported kernel vanishing outside of  $[-1, 1]$  and satisfy  $\int_{\mathbb{R}} zK(z) dz = 0$  and  $\eta = \int_{\mathbb{R}} z^2 K(z) dz > 0$ ;
3.  $f_j^{u_j}$  is bounded and can be continuously differentiated twice at any given  $q^{u_j}$ ;
4.  $\mu_j \in (0, f_j^{u_j})$  at (5);
5.  $f^{\mathbf{u}}$  is also bounded and twice continuously differentiable and  $f_{(j,j)}^{\mathbf{u}}(q^{\mathbf{u}}) = \frac{\partial^2}{\partial (q^{u_j})^2} f^{\mathbf{u}}(q^{\mathbf{u}})$  is bounded and Lipchitz continuous;
6. Let  $\alpha = 1$  and  $C_\alpha = 1$  as  $n \rightarrow \infty$ . Also, there exists  $\theta > 0$ ,  $\beta \in (0, \min(6\theta, 1))$  and  $C_\beta = \lim_{n \rightarrow \infty} n^{(\alpha+\beta)/2-1} b_n^{-d/2} \geq 0$  such that  $b_n \sim n^{-\theta}$  as  $n \rightarrow \infty$ , and  $\beta < 2C_\alpha$ .

If these are held a sequence

$$n^{\beta/2} \left[ f_n^{\mathbf{u}} - f^{\mathbf{u}} - \frac{\eta B_n}{2} \sum_{j=1}^d f_{(j,j)}^{\mathbf{u}} \right]$$

where  $B_n = b_n^2(1 - 2\theta)^{-1}(1 + o(1))$  is asymptotically normally distributed with mean 0 and variance  $C_\beta^2 \tilde{\eta}^d f^{\mathbf{u}}(2 - \beta)^{-1}$  given  $\tilde{\eta} = \int_{\mathbb{R}} K^2(z) dz$ .

For the proof of a more general version of the theorem a reader is referred to the draft of Camirand and Decrouez, (2018).

**Remark 1.** It follows from the sixth assumption in Theorem 3 that  $\beta - \alpha + d\theta \leq 0$ . Therefore,  $\beta = \alpha - d\theta$  guarantees the optimal rate of convergence with convergence rate is given by  $\sqrt{n^{1-d\theta}} \sim \sqrt{nb_n^d}$  which in the context of kernel density estimation appears to be a usual non-parametric rate of convergence.

**Remark 2.** The application of this theorem lies not only in the estimation of the joint density but it can also be applied for the estimation of the marginals. It can be simply achieved just by taking  $d = 1$ .

Second, let us state the asymptotic squared error of the joint density estimator  $f_n^{\mathbf{u}}$

**Theorem 4** (Asymptotic squared error of the joint density estimator). *When  $n$  is large*

$$\mathbb{E}(f_n^{\mathbf{u}}) - f^{\mathbf{u}} = \frac{\eta B_n}{2} \sum_{j=1}^d f_{(j,j)}^{\mathbf{u}},$$

where  $B_n = b_n^2(1 - 2\theta)^{-1}(1 + o(1))$ .



A proof can be found in Camirand and Decrouez, (2018). The authors provide a more general proof of the Theorems 3 and 4.

Once we have an algorithm to estimate both the joint distribution and product of marginal distributions with almost sure convergence described in Theorem 3, we now may present, to our best knowledge, for the first time the recursive non-parametric estimator for the copula density function

$$c_n = \frac{f_n^{\mathbf{u}}}{f_{1,n}^{u_1} \cdots f_{d,n}^{u_d}}. \quad (8)$$

Let us introduce the asymptotic convergence for the copula density estimator presented in (8)

**Theorem 5.** *Consider the following assumptions*

1.  $K$  is a Lipchitz continuous twice differentiable kernel such that  $\int_{\mathbb{R}} K(z) dz = 1$ ,  $\int_{\mathbb{R}} zK(z) dz = 0$ , and  $\eta = \int_{\mathbb{R}} z^2 K(z) dz \in (0, \infty)$ ;
2.  $f^{\mathbf{u}}$  is bounded and twice continuously with respect to  $j$ -th component;
3. Given bandwidth  $b_{j,n} \rightarrow 0$  and  $nb_{j,n} \rightarrow \infty$  assumed the following: (i)  $n^{\beta/2}(b_{1,n}^2, \dots, b_{d,n}^2) \rightarrow (c_1, \dots, c_d)$ ; (ii)  $n^{1-\beta} \prod_{j=1}^d b_{j,n} \rightarrow k$  as  $n \rightarrow \infty$ ;
4.  $c_1, \dots, c_d \geq 0$ ,  $k > 0$ , and  $\beta \in (0, 1)$ .

If these assumptions are held, then

$$\left( n \prod_{j=1}^d b_{j,n} \right)^{1/2} \left[ c_n - c - \frac{2 \cdot \mathcal{B}(q^{\mathbf{u}})}{k \cdot (2 - \beta)} \right],$$

where

$$\mathcal{B}(q^{\mathbf{u}}) := \tilde{B}(q^{\mathbf{u}}) \prod_{j=1}^d \frac{1}{f_j^{u_j}(q^{u_j})} - c \sum_{j=1}^d \frac{B_j(q^{u_j})}{f_j^{u_j}(q^{u_j})}$$

given

$$\tilde{B}(q^{\mathbf{u}}) := \frac{\eta}{2} \sum_{j=1}^d c_j f_{(j,j)}^{\mathbf{u}}$$

and

$$B_j^{u_j} := \frac{\eta}{2} c_j f_j^{u_j}(q^{u_j})$$

is asymptotically normally distributed with mean 0 and variance

$$\frac{\nu^d c^2}{k^{3/2} (2 - \beta) f^{\mathbf{u}}}.$$

**Remark 3.** Note that the copula density estimator is asymptotically unbiased in case all  $c_j = 0$ ,  $j = 1 \dots d$

## 5 Bandwidth Selection

The notion of kernels comes with the concept of a bandwidth or a window width. A smaller value of the bandwidth the less the approximated function is smoothed. In this paper, there are three different equations where a bandwidth appears.

First of all, the estimation of marginals  $f_{j,n}^{u_j}$  involves kernels with bandwidths  $h_{j,i}$  (see (4)). Another set of bandwidths appear in the estimation of the joint density  $f_n^{\mathbf{u}}$  as we propose to use the product of kernels  $W_t^{\mathbf{u}}$  with bandwidth  $b_{j,n}$  to estimate it (see (7)).

In this section, we present several ways of how these bandwidths might be calculated.

### 5.1 Data-independent Bandwidths

The first idea is to use only one data-independent (constant) bandwidth within estimation of all quantities. This type of bandwidths satisfies the required assumptions and quite easy to understand. A family of such bandwidths can be defined in the following way

$$h_{j,n} = n^{-\tau},$$

where  $\tau \in (0, 1/2)$  as it was proposed in Robinson, (1975) and Holst, (1987) for non-smoothed recursive algorithms (without kernels). It is eligible to use this family of bandwidths satisfy the corresponding assumptions of Theorems 1 and 2.

### 5.2 Silverman's Bandwidths

The first data-based bandwidth is the Silverman's "rule-of-thumb" bandwidth. The "Silverman's bandwidth" uses aggregated information gathered from the data, i.e. the mean and standard deviation. The definition of the bandwidth for a univariate case when kernel  $K$  is standard normal kernel is as follows (Silverman, 1986)

$$h_n = \left( \frac{4}{3} \sigma^5 \right)^{\frac{1}{5}} \cdot n^{-\frac{1}{5}}.$$

A recursive version of this expression involves a recursive estimation of sample standard deviation  $S_{j,n}^2$  as well as the sample mean  $\bar{X}_{j,n}$ . Consider

$$\begin{aligned} h_{j,n} &= \left( \frac{4}{3} S_{j,n}^5 \right)^{\frac{1}{5}} \cdot n^{-\frac{1}{5}} \\ S_{j,n}^2 &= \frac{n-2}{n-1} S_{j,n-1}^2 + \frac{(X_{j,n} - \bar{X}_{j,n-1})^2}{n} \\ \bar{X}_{j,n} &= \bar{X}_{j,n-1} + \frac{X_{j,n} - \bar{X}_{j,n-1}}{n}. \end{aligned}$$

A more detailed overview of the method could be found in the mentioned paper.

### 5.3 Specialized Bandwidths

The idea of using an individual bandwidth for both quantities  $b_j$ ,  $h_j$  (in  $f_j^{u_j}$ , and  $W_b^{\mathbf{u}}$  respectively) is quite natural. Therefore we present data-based recursive estimation of bandwidths for the marginals  $f_j^{u_j}$  and for the joint density  $f^{\mathbf{u}}$ .

So, (8) has the joint distribution and marginals that we want to estimate. We will start the derivation of the optimal bandwidth for the joint distribution  $f^{\mathbf{u}}$  because it is a general case and if it is possible to derive the optimal bandwidth, then it is also applicable for the marginals which are just a special case.

Now we may easily take the formulas for bias and variance of the joint density estimator  $f_n^{\mathbf{u}}$  from Theorem 3

$$\begin{aligned}\text{Bias}(f_n^{\mathbf{u}}, f^{\mathbf{u}}) &= \frac{\eta b_n^2}{2(1-2\theta)} \sum_{j=1}^d f_{(j,j)}^{\mathbf{u}} (1 + o(1)) \\ \text{Variance}(f_n^{\mathbf{u}}) &= \frac{C_\beta^2 \tilde{\eta}^d f^{\mathbf{u}}}{n^\beta (2-\beta)} (1 + o(1))\end{aligned}$$

and calculate the mean squared error between the true  $f^{\mathbf{u}}$  and our estimator  $f_n^{\mathbf{u}}$ . So, it can be presented in the following way

$$\begin{aligned}\text{MSE}(f_n^{\mathbf{u}}) &= \text{Bias}^2(f_n^{\mathbf{u}}, f^{\mathbf{u}}) + \text{Variance}(f_n^{\mathbf{u}}) \\ &= b_n^4 \frac{\eta^2 \cdot (\sum_{j=1}^d f_{(j,j)}^{\mathbf{u}})^2}{4(1-2\theta)^2} + \frac{C_\beta^2 \tilde{\eta}^d f^{\mathbf{u}}}{n^\beta \cdot (2-\beta)}.\end{aligned}$$

Then, we may optimize on  $b_n$  and obtain

$$b_n = \arg \min_{b_n} \{\text{MSE}(f_n^{\mathbf{u}})\} = \left[ \frac{\tilde{\eta}^d \cdot f^{\mathbf{u}} \cdot (d+2) \cdot d}{2 \cdot \eta^2 \cdot [\sum_{j=1}^d f_{(j,j)}^{\mathbf{u}}]^2 \cdot (d+4)} \right]^{\frac{1}{d+4}} \cdot n^{-\frac{1}{d+4}}. \quad (9)$$

For the proof of (9) a reader is refereed to Appendix B. This result is consistent with Mokkadem et al., (2009) as well. The presented bandwidth  $b_n$  is used in order to estimate the joint density introduced in (7).

## 6 Numerical Experiments

This section presents results for several numerical studies that show converges of copula density estimator introduced in (7) and (8). An experiment set up can be defined as follows:

1. Select a copula which properties and analytical expression for the density are known;
2. Draw a sample  $\mathbf{X} \in \mathbb{R}^{N \times d}$  of points from a selected copula;
3. Select a grid of quantiles  $\mathbf{u} \in (0, 1)^{k \times k}$  (for  $d = 2$ ) that we want to estimate. In this paper, we use nine quantiles  $(0.1, \dots, 0.9)$ , space of points is  $9 \times 9$ ,  $d = 2$ ;

4. Estimate the copula density at each point of the grid  $\mathbf{u}$  using (8);
5. Calculate the mean squared error between the ground truth and estimated copula density values using the proposed algorithm at each quantile  $\mathbf{u}$ .

To show the robustness of the algorithms we run the experiment for both types of copulas: Archimedean (Frank and Clayton copula) and Elliptical (Student copula) with different parameters  $\theta$ . (The mathematical definition of most popular copulas are provided in Appendix A, on the page 17). In addition to different copula types, the results are also provided for different choices of bandwidth.

We use  $\theta = \{-1, 0, 1\}$  for the Frank copula,  $\theta = \{-0.5, 0.5, 10\}$  for the Clayton copula, and  $\theta = \{-0.5, 0, 0.5\}$  for the Student copula. The shapes of these copulas for these parameters are presented in Figure 2 on the page 20.

The rationale behind this choice of parameters and copula types is the diversity of dependence structures that these copulas may capture. The goal is to choose not only some easy cases like Frank copulas with  $\theta = \{-1, 0, 1\}$  but also some hard cases like Clayton with  $\theta = -0.5$  where there is a discontinuity and Clayton with  $\theta = 10$  where there is a high spike in one corner (see also Figure 2).

In addition, the numerical experiments are run for the three different sets of bandwidths as it is presented in Section 5. It is expected that data-based approaches should perform better at least in the hardest cases like Clayton copula  $\theta = 10$ . All experiments are run for 100 Monte-Carlo simulations with 5000 i.i.d. bivariate draws.

The numerical results that were received by following the set-up rules presented earlier are depicted on the Figure 1. It can be seen that depending on how complicated a copula is the performance of the algorithms that use data-dependent bandwidths changes dramatically. Compare the results of the Silverman’s bandwidth on the plot of Student copula with  $\theta = -0.5$  with the results it achieves in more uncomplicated cases like the results of Frank copula.

Besides, even though the algorithm that uses the data-independent bandwidth appears to be quite shallow performs surprisingly well but only in simple cases like in Frank copula for all  $\theta$  that was considered. Therefore, it might be used as a rule-of-thumb bandwidth along with a historically accepted Silverman’s.

Finally, in case of the algorithms that rely on the specialized bandwidth we may admit the obvious convergence as well as the best performance in the most complicated case of Clayton copula with  $\theta = 10$ . However, it seems to be outperformed by the more straightforward candidates. Therefore, more parameter tuning is required to achieve better performance.

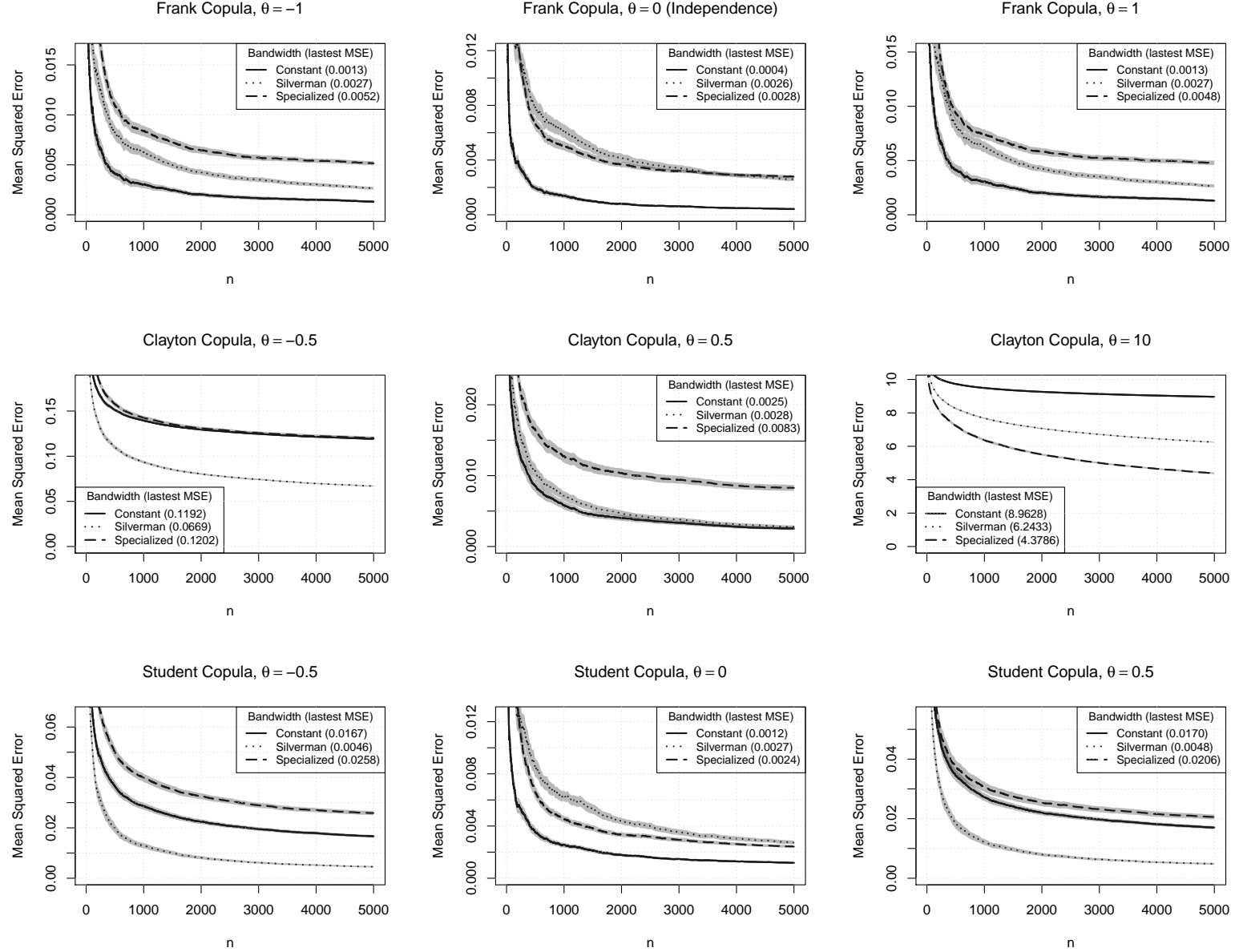


Figure 1: Goodness of fit in numerical experiments for different types of copulas and different choices of bandwidth ( $\#$  of samples = 5K,  $\#$  of Monte Carlo simulations = 100, 95 % confidence interval).

## 7 Conclusion

The contribution of this paper is two-fold. First of all, we introduced an algorithm of estimation of the multivariate joint density function using the smoothed non-parametric recursive estimator. Secondly, we successfully applied this estimator to the concept of copulas, joint distributions that enable to capture complex dependence structures among its variates.

The performance of the introduced algorithm for different experiment settings is successfully shown in numerical experiments. It is seen when a copula describes a reasonably complex, spiky, dependence structure the algorithms that use data-based bandwidths perform better, and, otherwise, data independent bandwidths achieve better results in more straightforward cases. The results of the experiments as well as other plots can be reproduced using the code that is available on <https://github.com/vdyashin/CopulaDensityEstimator>.

There is still research should be done in the area of automatic bandwidth selection as in most of the cases a simple data-independent bandwidth outperforms the more complex ones. In addition, it is also essential to find an application to this powerful analytical tool that is able to handle real-time parameter updates. We inclined to believe that the presented algorithms will be quite handy in estimating the dependence structure in conditions where the data-flow is ongoing.

## Acknowledgements

The author of this master thesis is grateful to his university Higher School of Economics for the provided opportunity to be an exchange student and be able to spend one semester in another campus where his scientific supervisor works and lives. The author also thanks the university stuff including office managers Maria Neklyudova as well as Ekaterina Pavlova and Ilona Yakovleva for their support and responsiveness. Special gratitude to Felix Camirand (Ph.D.) From University of Melbourne who has provided an early version of his paper.

I would also like to thank the uncountable number of contributors of **R** package and the creators of the **R** library `copula`. The contributors of the L<sup>A</sup>T<sub>E</sub>X typesetting system are also gratefully acknowledged.

## References

- Amiri, A. and B. Thiam (2014). “A smoothing stochastic algorithm for quantile estimation”. In: *Statistics & Probability Letters* 93, pp. 116–125. ISSN: 0167-7152.
- Andersen, L. and J. Sidenius (2004). “Extensions to the Gaussian copula: Random recovery and random factor loadings”. In: *Journal of Credit Risk Volume* 1.1, p. 5.
- Camirand, F. and G. Decrouez (2018). “(Draft) Non-parametric recursive estimation of the copula”.
- Cherubini, U., E. Luciano, and W. Vecchiato (2004). *Copula methods in finance*. John Wiley & Sons. ISBN: 0470863455.

- Chu, B. (2011). “Recovering copulas from limited information and an application to asset allocation”. In: *Journal of Banking & Finance* 35.7, pp. 1824–1842. ISSN: 0378-4266.
- Davies, H. I. (1973). “Strong consistency of a sequential estimator of a probability density function”. In: *Bull. Math. Statist* 15, pp. 49–54.
- Deheuvels, P. (1978). “Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes”. In: *Publ. Inst. Statist. Univ. Paris* 23, pp. 1–36.
- (1979). “La fonction de dépendance empirique et ses propriétés. Académie Royale de Belgique”. In: *Bulletin de la Classe des Sciences* 65.5, pp. 274–292.
- Dempster, M. A. H., E. A. Medova, and S. W. Yang (2007). “Empirical copulas for CDO tranche pricing using relative entropy”. In: *International Journal of Theoretical and Applied Finance* 10.04, pp. 679–701. ISSN: 0219-0249.
- Dupuis, D. et al. (2009). “Empirical evidence on the dependence of credit default swaps and equity prices”. In: *Journal of Futures Markets* 29.8, pp. 695–712. ISSN: 1096-9934.
- Erhardt, V. and C. Czado (2010). “Sampling count variables with specified pearson correlation: A comparison between a naive and a c-vine sampling approach”. In: *Dependence modeling: Vine copula handbook*. World Scientific, pp. 73–87.
- (2012). “Modeling dependent yearly claim totals including zero claims in private health insurance”. In: *Scandinavian Actuarial Journal* 2, pp. 106–129. ISSN: 0346-1238.
- Favre, A-C. et al. (2004). “Multivariate hydrological frequency analysis using copulas”. In: *Water resources research* 40.1. ISSN: 1944-7973.
- Friedman, C. and J. Huang (2010). “Most entropic copulas: General form, and calibration to high-dimensional data in an important special case”. In:
- Genest, Ch. et al. (2007). “Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data”. In: *Water Resources Research* 43.9. ISSN: 1944-7973.
- Hofert, M. and Mächler, M. (2011). “Nested Archimedean Copulas Meet R: The nacopula Package”. In: *Journal of Statistical Software* 39.9, pp. 1–20. URL: <http://www.jstatsoft.org/v39/i09/>.
- Hofert, M et al. (2017). *copula: Multivariate Dependence with Copulas*. R package version 0.999-18. URL: <https://CRAN.R-project.org/package=copula>.
- Holst, U. (1987). “Recursive estimation of quantiles using recursive kernel density estimators”. In: *Sequential Analysis* 6.3, pp. 219–237. ISSN: 0747-4946.
- Hyrš, M. and J. Schwarz (2015). “Elliptical and archimedean copulas in estimation of distribution algorithm with model migration”. In: *Computational Intelligence (IJCCI), 2015 7th International Joint Conference on*. Vol. 1. IEEE, pp. 212–219. ISBN: 9897581650.
- Kojadinovic, I. and Yan, J. (2010). “Modeling Multivariate Distributions with Continuous Margins Using the copula R Package”. In: *Journal of Statistical Software* 34.9, pp. 1–20. URL: <http://www.jstatsoft.org/v34/i09/>.
- Kresta, A. and T. Tichý (2012). “International Equity Portfolio Risk Modeling: The Case of the NIG Model and Ordinary Copula Functions”. In: ISSN: 0015-1920.
- Li, D. X. (1999). “On default correlation: A copula function approach”.
- Matthias, S. and M. Jan-frederik (2017). *Simulating copulas: stochastic models, sampling algorithms, and applications*. Vol. 6. Singapore World Scientific Publishing Co. Pte Ltd., p. 357. ISBN: 9813149264.

- Mokkadem, A., M. Pelletier, and Y. Slaoui (2009). “The stochastic approximation method for the estimation of a multivariate probability density”. In: *Journal of Statistical Planning and Inference* 139.7, pp. 2459–2478. ISSN: 0378-3758.
- Moore, D. S. and M. C. Spruill (1975). “Unified large-sample theory of general chi-squared statistics for tests of fit”. In: *The Annals of Statistics*, pp. 599–616. ISSN: 0090-5364.
- Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media. ISBN: 0387286780.
- Patton, A. J. (2009). “Copula-based models for financial time series”. In: *Handbook of financial time series*. Springer, pp. 767–785.
- (2012). “A review of copula models for economic time series”. In: *Journal of Multivariate Analysis* 110, pp. 4–18. ISSN: 0047-259X.
- Piantadosi, J., P. G. Howlett, and J. Borwein (2012). “Copulas with maximum entropy”. In: *Optimization Letters* 6.1, pp. 99–125. ISSN: 1862-4472.
- Piantadosi, J., P. G. Howlett, J. Borwein, and J. Henstridge (2012). “Maximum entropy methods for generating simulated rainfall”. In: *Numer. Algebra Control Optim* 2, pp. 233–256.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Renard, B. and M. Lang (2007). “Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology”. In: *Advances in Water Resources* 30.4, pp. 897–912. ISSN: 0309-1708.
- Robbins, H. and S. Monro (1951). “A stochastic approximation method”. In: *The annals of mathematical statistics*, pp. 400–407. ISSN: 0003-4851.
- Robinson, D. W. (1975). “Non-parametric quantile estimation through stochastic approximation.” PhD thesis.
- Ruschendorf, L. (1976). “Asymptotic distributions of multivariate rank order statistics”. In: *The Annals of Statistics*, pp. 912–923. ISSN: 0090-5364.
- Salvadori, G. et al. (2007). *Extremes in nature: an approach using copulas*. Vol. 56. Springer Science & Business Media, p. 292. ISBN: 1402044151.
- Schönbucher, P. J. and D. Schubert (2001). “Copula-dependent default risk in intensity models”. Silva, P. P. da, P. T. Rebelo, and C. Afonso (2014). “Tail Dependence of Financial Stocks and CDS Markets-Evidence Using Copula Methods and Simulation-Based Inference”. In: *Economics* 8.39. ISSN: 1864-6042.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New-York: Routledge, p. 176. ISBN: 1351456172.
- Sklar, A. (1959). “Fonctions de repartition an dimensions et leurs marges”. In: *Publ. Inst. Statist. Univ. Paris* 8, pp. 229–231.
- Slaoui, Y. (2014). “The stochastic approximation method for estimation of a distribution function”. In: *Mathematical Methods of Statistics* 23.4, pp. 306–325. ISSN: 1066-5307.
- Strelen, J. C. (2009). “Tools for dependent simulation input with copulas”. In: *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), p. 30. ISBN: 9639799459.



- Strelen, J. C. and F. Nassaj (2007). “Analysis and generation of random vectors with copulas”. In: *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*. IEEE Press, pp. 488–496. ISBN: 1424413060.
- Wegman, E. J. and H. I. Davies (1979). “Remarks on some recursive estimators of a probability density”. In: *The Annals of Statistics*, pp. 316–327. ISSN: 0090-5364.
- Wolverton, C. and T. Wagner (1969). “Asymptotically optimal discriminant functions for pattern classification”. In: *IEEE Transactions on Information Theory* 15.2, pp. 258–265. ISSN: 0018-9448.
- Yamato, H. (1971). “Sequential estimation of a continuous probability density function and mode”. In: *Bull. Math. Statist* 14, pp. 1–12.
- Yan, J. (2007). “Enjoy the Joy of Copulas: With a Package copula”. In: *Journal of Statistical Software* 21.4, pp. 1–21. URL: <http://www.jstatsoft.org/v21/i04/>.

# A Types of Copulas

There are two main classes of copulas: *Archimedean* and *Elliptical*. The Archimedean copulas are well-known for its simplicity in derivation and capability of capturing dependence structure quite well. The Elliptical copulas differ from the Archimedean ones in a way that they have an implicit expression that is based on the elliptical distribution families. Here we will follow the notation of Following the notations of Hyřš and Schwarz, (2015).

## A.1 Archimedean copulas

The Archimedean copulas are defined through the *generator* function. A generator is a function  $\psi : [0, \infty) \rightarrow [0, 1]$  which has the following properties:

1.  $\psi(0) = 1$  and  $\lim_{t \rightarrow \infty} \psi(t) = 0$ ;
2.  $\psi(\cdot)$  is continuous;
3.  $\psi(\cdot)$  is decreasing on  $[0, \infty)$  and strictly decreasing on  $[0, \inf\{t > 0 : \psi(t) = 0\})$ , given  $\psi(\emptyset) := 0$

$d$ -dimensional Archimedean copula is defined as follows

$$C_\psi(u_1, \dots, u_d) := \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)),$$

where  $\psi(\cdot)$  denotes the generator and  $\psi^{-1}(\cdot)$  its inverse.

The rationale behind the choice of this type for empirical applications is that by definition it may capture a wide range of dependence structure using different generators. In addition, it is quite easy to derive and estimate.

Next we define the three most popular Archimedean copulas: Frank, Clayton, and Gumbel. Also, we provide with definition of Independence copula. All of them has a simple functional form and one parameter  $\theta$  that depicts the dependency strength.

### A.1.1 Independence Copula

The simplest copula is the independence copula

$$\begin{aligned} C(u_1, u_2) &= u_1 u_2 \\ \psi(t) &= -\log(t) \\ \psi^{-1}(t) &= -\exp[-t] \end{aligned}$$

### A.1.2 Frank Copula

The bivariate Frank copula is defined as follows

$$\begin{aligned} C(u_1, u_2) &= -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right) \\ \psi(t) &= -\frac{1}{\theta} \log (e^{-t}(e^{-\theta} - 1) + 1) \\ \psi^{-1}(t) &= -\log \left( \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right) \\ \theta &\in \mathbb{R} \setminus \{0\} \end{aligned}$$

### A.1.3 Gumbel Copula

We can define bivariate Gumbel copula

$$\begin{aligned} C(u_1, u_2) &= \exp \left[ - \left( (-\log(u_1))^\theta + (-\log(u_2))^\theta \right)^{1/\theta} \right] \\ \psi(t) &= \exp \left[ - t^{1/\theta} \right] \\ \psi^{-1}(t) &= (-\log(t))^\theta \\ \theta &\in [1, \infty) \end{aligned}$$

### A.1.4 Clayton Copula

And finally bivariate Clayton copula

$$\begin{aligned} C(u_1, u_2) &= \max [u_1^{-\theta} + u_2^{-\theta} - 1, 0]^{-1/\theta} \\ \psi(t) &= (1 + t)^{-1/\theta} \\ \psi^{-1}(t) &= t^{-\theta} - 1 \\ \theta &\in [-1, \infty) \setminus \{0\} \end{aligned}$$

## A.2 Elliptical Copulas

Compared with Archimedean copulas, elliptical copulas have only implicit formulation that follows from the Sklar's theorem. A  $d$ -variate elliptical copula is defined in the following fashion

$$C(u_1, u_2) := F_\Theta(F^{-1}(u_1), F^{-1}(u_2)),$$

where  $F_\Theta(\cdot)$  represents an elliptical distribution with covariance matrix  $\Theta$  and  $F^{-1}(\cdot)$  its inverse. The covariance matrix  $\Theta$  has  $(d^2 - d)/2$  parameters to express the dependence structure. Therefore, this class of copulas may capture a wider range of dependencies in comparison with the Archimedean class (Hyrš and Schwarz, 2015).

There are two most popular types of copulas in the elliptical class: Normal (Gaussian) and Student copulas.

### A.2.1 Student Copula

The bivariate Student copula that is governed by  $t$ -distribution and defined as follows

$$\begin{aligned} C(u_1, u_2) &= t_{\Theta, \eta}(t_{\eta}^{-1}(u_1), t_{\eta}^{-1}(u_2)) \\ &= \int_{-\infty}^{t_{\eta}^{-1}(u_1)} \int_{-\infty}^{t_{\eta}^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \left(1 + \frac{x^2 - 2\theta xy + y^2}{\eta(1-\theta^2)}\right)^{-(\eta+2)/2} dx dy \\ &\quad \theta \in (-1, 1) \end{aligned}$$

where  $t_{\Theta, \eta}$  is the multivariate  $t$ -distribution and  $t_{\eta}^{-1}(u_1)$  is the quantile of the univariate  $t$ -distribution with  $\eta$  degrees of freedom.

### A.2.2 Gaussian Copula

Similarly, the bivariate Normal (Gaussian) copula is governed by Normal distribution

$$\begin{aligned} C(u_1, u_2) &= \Phi_{\Theta}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left[-\frac{x^2 - 2\theta xy + y^2}{2(1-\theta^2)}\right] dx dy \\ &\quad \theta \in (-1, 1) \end{aligned}$$

The illustration of all of the described copulas can be seen in Figure 2. Also note that in most of the copulas there is exist a value of  $\theta$  such that the copula degenerate into independence copula (see. Frank and Normal copulas with  $\theta \approx 0$ ).

## B Calculation of Bandwidth for the Joint Distribution

Assume  $\alpha = 1$ , then  $C_{\alpha} = 1$ . A bandwidth can be presented as

$$b_n = C_h n^{-\theta} \tag{10}$$

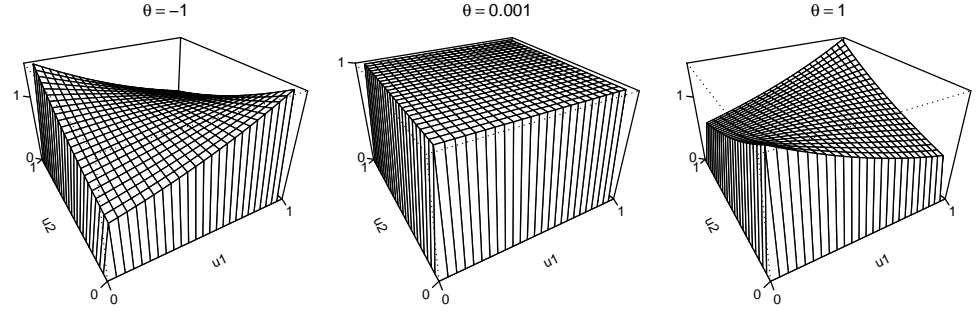
So, let us start with the bias and variance formulas that are obtained from Theorem 3

$$\begin{aligned} \text{Bias}(f_n^{\mathbf{u}}, f^{\mathbf{u}}) &= \frac{\eta b_n^2}{2(1-2\theta)} \sum_{j=1}^d f_{(j,j)}^{\mathbf{u}} (1 + o(1)) \\ \text{Variance}(f_n^{\mathbf{u}}) &= \frac{C_{\beta}^2 \tilde{\eta}^d f^{\mathbf{u}}}{n^{\beta}(2-\beta)} (1 + o(1)). \end{aligned}$$

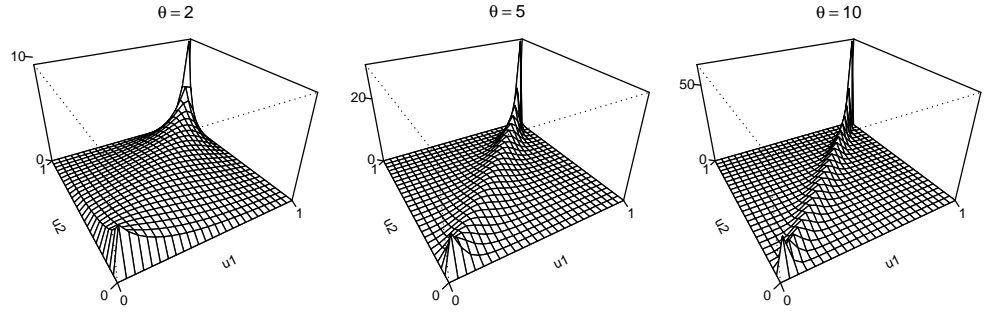
Then mean squared error is as follows

$$\begin{aligned} \text{MSE}(f_n^{\mathbf{u}}) &= \text{Bias}^2(f_n^{\mathbf{u}}, f^{\mathbf{u}}) + \text{Variance}(f_n^{\mathbf{u}}) \\ &= b_n^4 \frac{\eta^2 \cdot (\sum_{j=1}^d f_{(j,j)}^{\mathbf{u}})^2}{4(1-2\theta)^2} + n^{-\beta} \frac{C_{\beta}^2 \tilde{\eta}^d f^{\mathbf{u}}}{2-\beta}. \end{aligned} \tag{11}$$

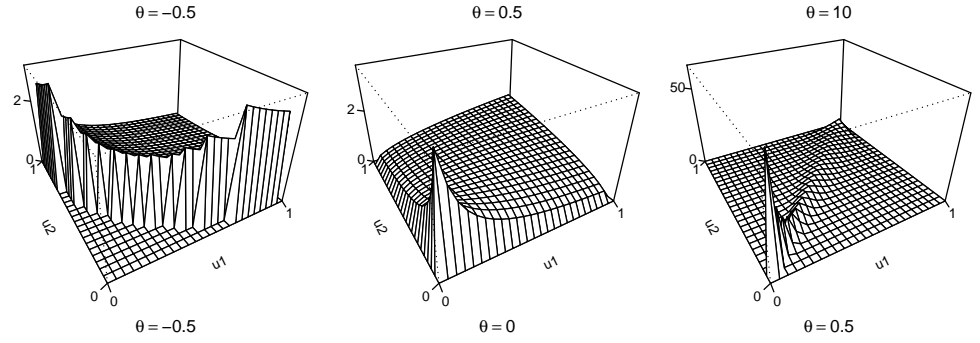
Frank Copula



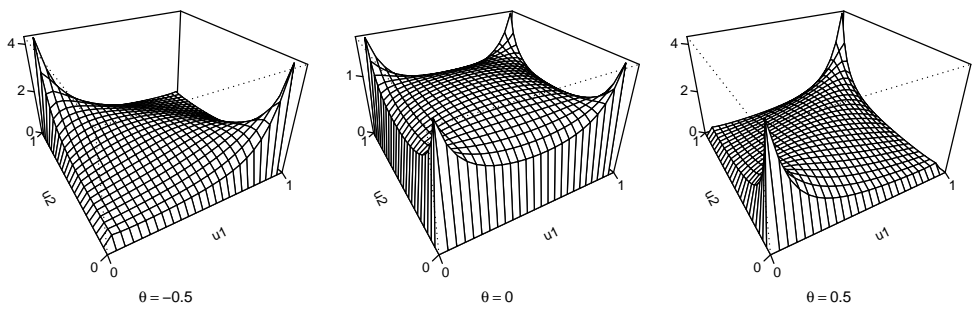
Gumbel Copula



Clayton Copula



Student Copula



Normal Copula

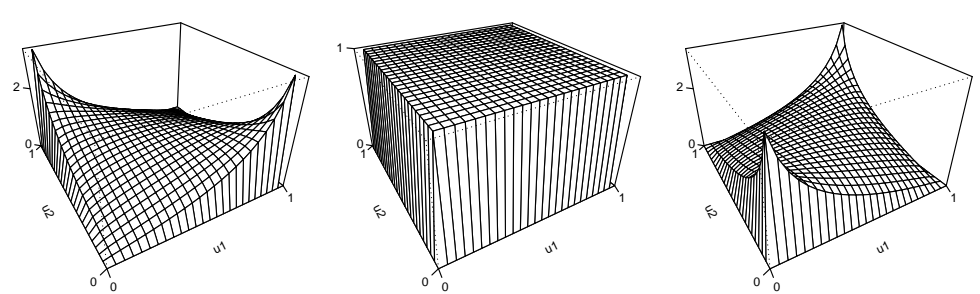


Figure 2: Different types of copulas with different parameter  $\theta$ .

Also, according to the Assumption 6 of Theorem 3 we put  $\beta = \alpha - d\theta = 1 - d\theta$ . Thus,  $\beta < 2$  and  $1 - d\theta < 2$ . Hence, the assumption is satisfied.

Now, let us rewrite  $n^{-\beta}$

$$n^{-\beta} = n^{-1+d\theta} = n^{-1}(n^{-\theta})^{-d},$$

re-express (10) as

$$\begin{aligned} n^{-\theta} &= b_n C_h^{-1} \rightarrow \\ n^{-\beta} &= n^{-1}(n^{-\theta})^{-d} = n^{-1}(b_n C_h^{-1})^{-d}. \end{aligned} \quad (12)$$

Then, we plug (12) into (11) and obtain the following

$$\begin{aligned} \text{MSE}(f_n^{\mathbf{u}}) &= b_n^4 \frac{\eta^2 \cdot (\sum_{j=1}^d f_{(j,j)}^{\mathbf{u}})^2}{4(1-2\theta)^2} + n^{-1}(b_n C_h^{-1})^{-d} \frac{C_\beta^2 \tilde{\eta}^d f^{\mathbf{u}}}{2-\beta} \\ &= b_n^4 \frac{\eta^2 \cdot (\sum_{j=1}^d f_{(j,j)}^{\mathbf{u}})^2}{4(1-2\theta)^2} + b_n^{-d} n^{-1} \frac{C_h C_\beta^2 \tilde{\eta}^d f^{\mathbf{u}}}{2-\beta} \\ &:= b_n^4 \cdot A_1 + b_n^{-d} n^{-1} \cdot A_2 \end{aligned}$$

After the optimization of  $\text{MSE}(f_n^{\mathbf{u}})$  on  $b_n$  we obtain

$$b_n^{4+d} = \frac{d \cdot A_2}{4n \cdot A_1}.$$

If we rewrite bandwidth in the form of (10) we get

$$b_n = \left[ \frac{d \cdot A_2}{4 \cdot A_1} \right]^{1/(d+4)} n^{-1/(d+4)}. \quad (13)$$

Thus, since  $b_n = C_h n^{-\theta}$ ,  $\theta = 1/(d+4)$ . Also, recall  $\beta = 1 - d\theta$ , then  $\beta = 4/(d+4)$ . We also know from the Assumption 6 of the Theorem 3 that  $\beta \in (0, \min\{6\theta, 1\})$ . This is satisfied if we plug  $\beta$  in the brackets.

Also note that

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{(\alpha+\beta)/2-1} h^{-d/2} &= C_\beta \\ &= n^{(1+\beta)/2-1} \cdot (C_h)^{-d/2}. \\ &= C_h^{-d/2} \cdot n^0 \\ &\Rightarrow C_h^{-d/2} = C_\beta. \end{aligned}$$

We plug  $\theta = 1/(d+4)$  into  $A_1$  and  $A_2$  which, in turn, we plug into (13) and obtain the bandwidth coefficient  $C_h$  from (10)

$$C_h = \left[ \frac{d \cdot C_h^d \cdot C_\beta^2 \cdot \tilde{\eta}^d \cdot f^{\mathbf{u} \cdot (d+2)}}{2 \cdot \eta^2 \cdot [\sum_{j=1}^d f_{(j,j)}^{\mathbf{u}}]^2 (d+4)} \right]^{1/(d+4)}.$$

We derived earlier that  $C_\beta = C_h^{-d/2}$ . Using this in the last formula we may obtain the form of bandwidth that is presented in Section 5.3.