Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466
Vol. 14, Issue 1 (June 2019), pp. 76–98

Applications and Applied Mathematics:
An International Journal
(AAM)

A Note on Asymptotic Normality of a Copula Function in Regression Model

[1]Khadidja Sabri, [2]Mustapha Meghnafi, [3]Abbes Rabhi and [4*]Djamila Bennafla

[1]Laboratory of Mathematics, University of Es-Senia Oran
P.O. BOX. 1524, Elm'Naouer 31000 Oran, Algeria
khadouj_2012@hotmail.com

[2]Faculty of Exact Sciences Department Maths
Tahri Mohamed University of Bechar
P.O. Box 417, 08000 Bechar, Algeria
megnafi3000@yahoo.fr

[3]Laboratory of Mathematics
University Djillali LIABES of Sidi Bel Abbés
PO. BOX. 89, Sidi Bel Abbés 22000, Algeria
rabhi_abbes@yahoo.fr

[4]Stochastic Models, Statistics and Applications Laboratory
Moulay Tahar University of Saida
P.O.Box 138 En-Nasr Saida 20 000, Saida, Algeria
dbennafla@hotmail.com

*Corresponding Author

Received: June 21, 2018; Accepted: March 22, 2019

Abstract

Over the last decade, there has been significant and rapid development of the theory of copulas. Much of the work has been motivated by their applications to stochastic processes, economics, risk management, finance, insurance, the environment (hydrology, climate, etc.), survival analysis, and medical sciences. In many statistical models. The copula approach is a way to solve the difficult problem of finding the whole bivariate or multivariate distribution. In this paper, we give the asymptotic normality of the copulas function in a regression model.

Keywords: Almost sure convergence; Asymptotic normality; Copulas; Functional Data Analysis; Kernel Estimator; Non parametric estimation; Regression model

MSC 2010 No.: 62N02, 62H12, 62J05

1. Introduction

The study of copulas is a growing field. The construction and properties of copulas have been studied rather extensively during the last 15 years or so. Hutchinson et al. (1990) were among the early authors who popularized the study of copulas. Nelsen (1999) presented a comprehensive treatment of bivariate copulas, while Joe (1997) devoted a chapter of his book to multivariate copulas. Further authoritative updates on copulas are given in Nelsen (2006). Copula methods have many important applications in insurance and finance Cherubini et al. (2004) and Embrechts et al. (2003).

Briefly speaking, copulas are functions that join or "couple" multivariate distributions to their one-dimensional marginal distribution functions. Equivalently, copulas are multivariate distributions whose marginals are uniform on the interval $(0, 1)$. In this chapter, we restrict our attention to bivariate copulas. Fisher (1997) gave two major reasons as to why copulas are of interest to statisticians: "Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions." Specifically, copulas are an important part of the study of dependence between two variables since they allow us to separate the effect of dependence from the effects of the marginal distributions. This feature is analogous to the bivariate normal distribution where the mean vectors are unlinked to the covariance matrix and jointly determine the distribution. Many authors have studied constructions of bivariate distributions with given marginals: this may be viewed as constructing a copula.

There is a fast-growing industry for copulas. They have useful applications in econometrics, risk management, finance, insurance, etc. The commercial statistics software SPLUS provides a module in fin metrics that include copula fitting written by Carmona (2004). One can also get copula modules in other major software packages such as R, Mathematica, Matlab, etc. The International Actuarial Association (2004) in a paper on Solvency II, recommends using copulas for modeling dependence in insurance portfolios. Moody's uses a Gaussian copula for modeling credit risk and provides software for it that is used by many financial institutions. Basle II2 copulas are now standard tools in credit risk management. There are many other applications of copulas, especially the Gaussian copula, the extreme-value copulas, and the Archimedean copula. We now classify these applications into several categories.

Nonparametric estimators of copula densities have been suggested by Gijbels et al. (1990) and Fermanian et al. (2005), who used kernel methods, Sancetta (2003) and Sancetta et al. (2004), who used techniques based on Bernstein polynomials. Biau et al. (2006) proposed estimating the copula density through a minimum distance criterion. Faugeras (2008) in his thesis studied the quantile copula approach to conditional density estimation.

The aim of this paper is devoted to the estimation of a regression model via a copulae function, the rest of the paper is organized as follows; at first we will introduce the model, then to state our result, we will have to make some regularity assumptions on the kernels and the densities which, although far from being minimal, are somehow customary in kernel density estimation, so we give some notations and give some assumptions and define the sklar's theorem, the main result and its proof is given in the third part of this paper.

2. The model

Let $((X_i; Y_i); i = 1, \dots, n)$ be an independent identically distributed sample from real-valued random variables (X, Y) sitting on a given probability space. For predicting the response Y of the input variable X at a given location x, it is of great interest to estimate not only the conditional mean or regression function $\mathbb{E}(Y|X = x)$, but the full conditional density $f(y|x)$. Indeed, estimating the conditional density is much more informative, since it allows not only to recalculate from the density the conditional expected value $\mathbb{E}(Y|X)$, but also many other characteristics of the distribution such as the conditional variance. In particular, having knowledge of the general shape of the conditional density, is especially important for multi-modal or skewed densities, which often arise from nonlinear or non- Gaussian phenomena, where the expected value might be nowhere near a mode, i.e., the most likely value to appear.

A natural approach to estimate the conditional density $f(y|x)$ of Y given $X = x$ would be to exploit the identity

$$f(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}, \;\; f_X(x) \neq 0, \tag{1}$$

where f_{XY} and f_X denote the joint density of (X, Y) and X, respectively.

By introducing Parzen (1962), Pickands (1975) kernel estimators of these densities, namely,

$$\widehat{f}_{n,XY}(x, y) = \frac{1}{n} \sum_{i=1}^{n} K'_{h'}(X_i - x) K_h(Y_i - y),$$

$$\widehat{f}_{n,X}(x) = \frac{1}{n} \sum_{i=1}^{n} K'_{h'}(X_i - x),$$

where $K_h(\cdot) = 1/hK(\cdot/h)$ and $K'_{h'}(\cdot) = 1/h'K'(\cdot/h')$ are (rescaled) kernels with their associated sequence of bandwidth $h = h_n$ and $h' = h'_n$ going to zero as $n \to 1$, one can construct the quotient

$$\widehat{f}_n(y|x) = \frac{\widehat{f}_{n,XY}(x, y)}{\widehat{f}_{n,X}(x)},$$

and obtain an estimator of the conditional density.

Formally, Sklar's theorem below elucidates the role that copulas play in the relationship between bivariate distribution functions and their univariate marginals (see Sklar (1959)).

Theorem 2.1.

For any bivariate cumulative distribution function $F_{X,Y}$ on $\mathbb{R}^2$, with marginal cumulative distribution functions F of X and G of Y, there exists some function $C : [0, 1]^2 \to [0, 1]$, called the dependence or copula function, such as

$$F_{X,Y}(x,y) = C(F(x), G(y)), \ -\infty \leq x, y \leq +\infty. \tag{2}$$

If F and G are continuous, this representation is unique with respect to (F, G). The copula function C is itself a cumulative distribution function on $[0,1]^2$ with uniform marginals.

It is Sklar's Theorem (Sklar (1959)) that gives a representation of the bivariate c.d.f. as a function of each univariate c.d.f. In other words, the copula function captures the dependence structure among the components X and Y of the vector (X, Y), irrespectively of the marginal distribution F and G. Simply put, it allows to deal with the randomness of the dependence structure and the randomness of the marginal separately.

Copulas appear to be naturally linked with the quantile transform: in the case F and G are continuous, formula (2) is simply obtained by defining the copula function as $C(u,v) = F_{X,Y}(F^{-1}(u), G^{-1}(v))$, $0 \leq u \leq 1$, $0 \leq v \leq 1$. For more details regarding copulas and their properties, one can consult for example the book of Joe (1997). Copulas have witnessed a renewed interest in statistics, especially in finance, since the pioneering work of Rüschendorf (1976) and Deheuvels (1979), who introduced the empirical copula pro- cess. Weak convergence of the empirical copula process was investigated by Deheuvels et al. (1981), Van der Vaart et al. (1996), Fermanian et al. (2004). For the estimation of the copula density, refer to Gijbels et al. (1990), Fermanian (2005) and Fermanian et al. (2003).

From now on, we assume that the copula function $C(u,v)$ has a density $c(u,v)$ with respect to the Lebesgue measure on $[0,1]^2$ and that F and G are strictly increasing and differentiable with densities f and g. $C(u,v)$ and $c(u,v)$ are then the cumulative distribution function (c.d.f.) and density respectively of the transformed variables $(U,V) = (F(x), G(y))$. By differentiating formula (2), we get for the joint density,

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} = \frac{\partial^2 C(F(x); G(y))}{\partial F(x) \partial G(y)} \frac{\partial F(x)}{\partial x} \frac{\partial G(y)}{\partial y} = f(x) g(y) c(F(x), G(y)),$$

where $c(u,v) := \frac{\partial^2 C(u,v)}{\partial u \partial v}$ is the above mentioned copula density. Eventually, we can obtain the following explicit formula of the conditional density

$$f(y|x) = \frac{f_{XY}(x,y)}{f(x)} = g(y) c(F(x), G(y)), \ \ f(x) \neq 0. \tag{3}$$

Concerning the copula density $c(u,v)$, we noted that $c(u,v)$ is the joint density of the transformed variables $(U,V) := (F(x), G(y))$. Therefore, $c(u,v)$ can be estimated by the bivariate Parzen-Rosenblatt kernel type non parametric density (pseudo) estimator,

$$c_n(u,v) := \frac{1}{n h_n b_n} \sum_{i=1}^{n} K\left(\frac{u - U_i}{h_n}, \frac{v - V_i}{b_n}\right), \tag{4}$$

where K is a bivariate kernel and h_n, b_n its associated bandwidth. For simplicity, we restrict ourselves to product kernels, i.e., $K(u,v) = K_1(u) K_2(v)$ with the same bandwidths $h_n = b_n$.

Nonetheless, since F and G are unknown, the random variables (U_i, V_i) are not observable, i.e., c_n is not a true statistic. Therefore, we approximate the pseudo-sample (U_i, V_i), $i = 1, \ldots, n$ by its empirical counterpart $(F_n(X_i), G_n(Y_i))$, $i = 1, \ldots, n$. We therefore obtain a genuine estimator of $c(u, v)$.

$$\widehat{c}_n(u, v) := \frac{1}{nh_n^2} \sum_{i=1}^{n} K_1 \left(\frac{u - F_n(X_i)}{h_n} \right) K_2 \left(\frac{v - G_n(Y_i)}{b_n} \right), \tag{5}$$

the empirical distribution functions $F_n(x)$ and $G_n(y)$ for $F(x)$ and $G(y)$ respectively,

$$F_n(x) = \sum_{j=1}^{n} 1_{X_j \leq x} \text{ and } G_n(y) = \sum_{j=1}^{n} 1_{Y_j \leq y}.$$

Our estimated model is given as follows: the regression function $r(x)$ is estimated by a function $\widehat{r}_n(x)$

$$r(x) = \mathbb{E}(Y|X = x) = \int yf(y|x)dy = \int yg(y)c(F(x), G(y))dy = \mathbb{E}(Yc(F(x), G(y))).$$

This regression function $r(x)$ is estimated by a function $\widehat{r}_n(x) = \int y\widehat{f}_n(y|x)dy$.

To state our result, we will have to make some regularity assumptions on the kernels and the densities which, although far from being minimal, are somehow customary in kernel density estimation.

3. Notations and Assumptions

We note the ith moment of a generic kernel (possibly multivariate) K as

$$m_i(K) := \int u^i K(u)du,$$

and the $\mathbb{L}_p$ norm of a function h by $\|s\|_p := \int s^p$. We use the sign $\simeq$ to denote the order of the bandwidths. Set (u, v) fixed point in the interior of $supp(c)$. The support of the densitie function c is noted by $supp(c) = \overline{\{(u, v) \in \mathbb{R}^2; c(u, v) > 0\}}$, where $\bar{A}$ stands for the closure of a set A. Finally, $\mathcal{O}_{\mathbb{P}}(\cdot)$ and $\mathcal{O}_p(\cdot)$ (respectively $o_{a.s}(\cdot)$ and $\mathcal{O}_{a.s}(\cdot)$) will stands for convergence and boundedness in probability (respectively almost surely).

3.1. Assumptions

(1) The c.d.f F of x and G of Y are strictly increasing and differentiable.
(2) The density c is twice continuously differentiable with bounded second derivatives on its support.
(3) The density c is uniformly continuous and non-vanishing almost everywhere on a compact set $D \subset (0, 1) \times (0, 1)$ included in the interior of $supp(c)$.
(4) K is of bounded support and of bounded variation.
(5) $0 \leq K \leq \alpha$ and $0 \leq K_0 \leq \alpha$ for some constant α.
(6) K is second order kernels, $m_0(K) = 1, m_1(K) = 0$ and $m_2(K) < +\infty$.

(7) K it is twice differentiable with bounded second partial derivatives.

Recall that $c_n(u,v)$ is the kernel copula (pseudo) density estimator from the unobservable, but fixed with respect to n, pseudo data $(F(X_i), G(Y_i))$, and that $\widehat{c}_n(u,v)$ is its analogue made from the approximate data $(F_n(X_i), G_n(Y_i)$. The heuristic of the reason why our estimator works is that the $n^{-1/2}$ in probability rate of convergence in uniform norm of F_n and G_n to F and G is faster than the $1/\sqrt{na_n^2}$ rate of the non parametric kernel estimator c_n of the copula density c. Therefore, the approximation step of the unknown transformations F and G by their empirical counterparts F_n and G_n does not have any impact asymptotically on the estimation step of c by c_n. Put in another way, one can approximate $\widehat{c}_n(F_n(x), G_n(y))$ by $c_n(F(x), G(y))$ at a faster rate than the convergence rate of $c_n(F(x), G(y))$ to $c(F(x), G(y))$.

3.2. Main Result

This part of the paper is devoted to the asymptotic study the convergence in probability and almost surely of our estimators introduced above. But at first let us present the rate convergence of the estimator.

Theorem 3.1.

Let the regularity assumptions (1)-(7) on the densitie and the kernel be satisfied, if h_n tends to zero as $n \to \infty$ in such a way that

$$nh_n^4 \to \infty, \frac{\sqrt{\ln \ln n}}{nh_n^3} \to 0,$$

then

$$\widehat{r}_n(x) = r(x) + \mathcal{O}_{\mathbb{P}}\left(h_n^2 + \frac{1}{\sqrt{nh_n^2}} + \frac{1}{nh_n^4} + \frac{\sqrt{\ln \ln n}}{nh_n^3}\right)$$

The main ingredient of the proof follows from the fact that

$$\widehat{r}_n(x) - r(x) = Y(\widehat{c}_n(F_n(x), G_n(y)) - c_n(F(x), G(y))).$$

On the one hand, convergence results for the kernel density estimators of what will follow entail that,

$$c_n(F(x), G(y)) - c(F(x), G(y)) = \mathcal{O}_{\mathbb{P}}(h_n^2 + 1/\sqrt{nh_n^2}).$$

Thus, by lemmas Lemma 3.13 of Faugeras (2008) and Lemma 3.5 respectively.

Corollary 3.2.

we get the rate of convergence, by choosing the bandwidth which balance the bias and variance trade-off: for an optimal choice of $h_n \simeq n^{-1/6}$, we get

$$\widehat{r}_n(x) = r(x) + \mathcal{O}_{\mathbb{P}}(n^{-1/3})$$

Therefore, our estimator is rate optimal in the sense that it reaches the minimax rate $n^{-1/3}$ of convergence.

Almost sure results can be proved in the same way: we have the following strong consistency result,

Theorem 3.3.

Let the regularity assumptions (1)-(7) on the densitie and the kernel be satisfied.If the bandwidth h_n tends to zero as $n \to \infty$ in such a way that

$$\frac{\sqrt{\ln n \ln\ln n}}{nh_n^3} \to 0, \frac{\ln\ln n}{nh_n^4} \to 0,$$

then

$$\widehat{r}_n(x) = r(x) + \mathcal{O}_{a.s}\left(h_n^2 + \sqrt{\frac{\ln\ln n}{nh_n^2}} + \frac{\ln\ln n}{nh_n^4} + \frac{\sqrt{\ln n \ln\ln n}}{nh_n^3}\right)$$

Proof:

For the proof of this theorem, It is sufficient to follow the same lines as the preceding theorem , but uses the a.s. results of the consistency of the kernel density estimators of lemmas Lemma 3.13 of Faugeras (2008) and Lemma 3.5 and of the approximation propositions Proposition 3.6 and Proposition 3.7. It is therefore similar and omitted. ■

Corollary 3.4.

For $h_n \simeq (\ln\ln n/n)^{1/6}$ which is the optimal trade-off between the bias and the stochastic term, one gets the optimal rate

$$\widehat{r}_n(x) = r(x) + \mathcal{O}_{a.s}\left(\frac{\ln\ln n}{n}\right)^{1/3}$$

Let $\widehat{r}_n(x) = Y\widehat{c}_n(F_n(x), G_n(x))$, to demonstrate that $\widehat{r}_n(x)$ converge to $r(x)$ it is sufficient to prove that $\widehat{c}_n(u, v) \to c_n(u, v)$.

Recall a preliminary result that will be needed for the main result.

For $(X_i, i = 1, \ldots, n)$ an i.i.d. sample of a real random variable X with common c.d.f. F, the Kolmogorov-Smirnov statistic is defined as $D_n := \|F_n - F\|_\infty$. Glivenko-Cantelli, Kolmogorov and Smirnov, Chung, Donsker among others have studied its convergence properties in increasing generality (See e.g. Shorack et al. (1986) and Van der Vaart et al. (1996) for recent accounts). For our purpose, we only need to formulate these results given in the Lemma 3.13 Faugeras (2008):

$$|F(X_i) - F_n(X_i)| \leq \sup_{x\in\mathbb{R}} |F(x) - F_n(x)| = \|F_n - F\|_\infty \quad a.s.$$

Apply this result to the estimator c_n, and for this let us introduce the following result Lemma 3.15 page 82 given in Faugeras (2008).

Lemma 3.5.

With the previous assumptions, for $(u,v) \in (0,1)^2$, we have,

(1) For a bandwidth chosen such as $h_n \simeq n^{-1/6}$,

$$|c_n(u,v) - c(u,v)| = \mathcal{O}_{\mathbb{P}}(n^{-1/3}).$$

(2) For a point (u,v) where $c(u,v) > 0$, and $h_n = o(n^{-1/6})$,

$$\sqrt{nh_n^2}\left(\frac{c_n(u,v) - c(u,v)}{\sqrt{c_n(u,v)\|K\|_2^2}}\right) \rightsquigarrow \mathcal{N}(0,1).$$

(3) For a bandwidth chosen of $h_n \simeq (\ln\ln n/n)^{1/6}$,

$$|c_n(u,v) - c(u,v)| = \mathcal{O}_{a.s}\left(\left(\frac{\ln\ln n}{n}\right)^{1/3}\right).$$

Proof:

The proof is given in Faugeras (2008). ■

Now, we need two proposed approximation $\widehat{c}_n$ by c_n.

Proposition 3.6.

Let $(u,v) \in (0,1)^2$. If the kernel $K(u,v) = K_1(u)K_2(v)$ is twice differentiable with bounded second derivative, then

$$|\widehat{c}_n(u,v) - c_n(u,v)| = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} + \frac{\sqrt{\ln\ln n}}{nh_n^3} + \frac{1}{nh_n^4}\right),$$

$$|\widehat{c}_n(u,v) - c_n(u,v)| = \mathcal{O}_{a.s}\left(\sqrt{\frac{\ln\ln n}{n}} + \frac{\sqrt{\ln n}\sqrt{\ln\ln n}}{nh_n^3} + \frac{\ln\ln n}{nh_n^4}\right).$$

Proof:

The proof is given in Faugeras (2008). ■

Proposition 3.7.

With the same assumptions as in the previous proposal was

(1) If $h_n \to 0, nh_n^3 \to \infty$

$$|\widehat{c}_n(F_n(x), G_n(y)) - c_n(F(x), G(y))| = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} + \frac{1}{nh_n^4}\right).$$

(2) If $h_n \to 0, nh_n^3/\ln\ln n \to \infty$

$$|\widehat{c}_n(F_n(x), G_n(y)) - c_n(F(x), G(y))| = \mathcal{O}_{a.s}\left(\sqrt{\frac{\ln\ln n}{n}} + \frac{\ln\ln n}{nh_n^4}\right).$$

Proof:

The proof is given in Faugeras (2008). ∎

4. Asymptotic Bias, Variance and Mean square error

When we want to work out the expectation and variance of the Nadaraya- Watson estimator, we face the difficulty that the numerator and denominator of this statistic are both random variables. Thus, let us do the analysis for the numerator and denominator separately. Define

$$r(x) = \int yf(y|x)dy = \int yf(x, y)dy = m(x)f(x).$$

Indeed, let's recall a well-known connection between the double-kernel estimator of the conditional density and the Nadaraya-Watson estimator of the regression function. Remember that the double kernel estimator of the conditional density writes

$$\widehat{f}(y|x) = \frac{\sum_{i=1}^{n} h_n^{-2} K\left(\frac{x-X_i}{h_n}\right) K(\frac{y-Y_i}{h_n})}{\sum_{i=1}^{n} h_n^{-1} K\left(\frac{x-X_i}{h_n}\right)}.$$

A natural plug-in approach to estimate the regression function $r(x)$ would be to integrate the double kernel estimator and define $\widehat{r}(x)$ as

$$\widehat{r}(x) = \int y\widehat{f}(y|x)dy = \frac{\sum_{i=1}^{n} h_n^{-2} K\left(\frac{x-X_i}{h_n}\right) \int_{i\infty}^{+\infty} yK\left(\frac{y-Y_i}{h_n}\right) dy}{\sum_{i=1}^{n} h_n^{-1} K\left(\frac{x-X_i}{h_n}\right)}.$$

By the change of variable formula, the above integral is equal to

$$h_n \int_{-\infty}^{+\infty} (Y_i + th_n)K(t)dt = h_n Y_i,$$

by the properties of a symmetric kernel K. Therefore, the plug-in estimator reduces to

$$\widehat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)},$$

which is the classical Nadaraya-Watson estimator of the regression function.

The regression curve estimate is thus given by

$$\widehat{m}_n(x) = \frac{\widehat{r}_n(x)}{\widehat{f}_n(x)},$$

with

$$\widehat{r}_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right).$$

$$\begin{aligned}
\mathbb{E}(\widehat{r}_n(x)) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right)\right] \\
&= \mathbb{E}\left[YK\left(\frac{x - X}{h_n}\right)\right] \\
&= \int\int yK\left(\frac{x-u}{h_n}\right) f(y|u)f(u)dydu \\
&= \int K\left(\frac{x-u}{h_n}\right) f(u)\int\left(yf(y|u)dy\right)du \\
&= \int K\left(\frac{x-u}{h_n}\right) f(u)(\mathbb{E}(Y|X=u))du \\
&= \int K\left(\frac{x-u}{h_n}\right) f(u)m(u)du \\
&= \int K\left(\frac{x-u}{h_n}\right) r(u)du. \qquad (6)
\end{aligned}$$

We note that if $r \in C^2$ and $h_n \to 0$

$$B_0 = \mathbb{E}(\widehat{r}_n(x)) - r(x) = \frac{h^2}{2} r^{''}(x)\mu_2(K) + o(h_n^2).$$

Hence, $r_n(x)$ is asymptotically unbiased for $(h_n \to 0)$.

To compute the variance of $r_n(x)$ let $S^2(x) = \mathbb{E}[y^2|X=x]$. Thus, for $nh_n \to \infty$, we have:

$$\begin{aligned}
Var(\widehat{r}_n(x)) &= Var\left[\frac{1}{n}\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right)\right] \\
&= \frac{1}{n} Var\left[YK\left(\frac{x-X}{h_n}\right)\right] \\
&= \frac{1}{n}\left[\int K^2\left(\frac{x-u}{h_n}\right)s^2(u)f(u)du - \left(\int K\left(\frac{x-u}{h_n}\right)r(u)du\right)^2\right] \\
&= \frac{1}{nh_n}\int K^2\left(\frac{x-u}{h_n}\right)s^2(x+uh)f(x+uh_n)du + o\left(\frac{1}{nh_n}\right) \\
&= = \frac{1}{nh_n} f(x)s^2(x)\|K\|_2^2 + o\left(\frac{1}{nh_n}\right). \qquad (7)
\end{aligned}$$

When we combine (6) and (7) we obtain the formula for the Mean squared error of $\widehat{r}_n(x)$ for $h \to 0, nh \to \infty$

$$MSE[\widehat{r}_n(x)] = \frac{1}{nh} f(x)s^2(x)\|K\|_2^2 + \frac{h_n^4}{4}(r^{''}(x)\mu_2(K))^2 + o(h_n^4) + o(\frac{1}{nh_n}).$$

Hence, if we let $h \to 0$ such that $nh \to \infty$, we have

$$MSE[\widehat{r}_n(x)] \to 0.$$

We have noted the difficulties in computing the MSE of $\widehat{m}(x)$ directly. In order to get more insight into how $\widehat{m}(x)$ behaves, we linearize the estimator as follows:

$$\begin{aligned}\widehat{m}(x) - m(x) &= \left(\frac{\widehat{r}_n(x)}{\widehat{f}_n(x)} - m(x)\right)\left(\frac{\widehat{f}_n(x)}{f(x)} + (1 - \frac{\widehat{f}_n(x)}{f(x)})\right) \\ &= \frac{\widehat{r}_n(x) - m(x)\widehat{f}_n(x)}{f(x)} + (\widehat{m}(x) - m(x))\frac{f(x) - \widehat{f}_n(x)}{f(x)}. \qquad (8)\end{aligned}$$

By the above consistency property of $\widehat{m}(x)$, we see that we can choose $h_n \sim n^{-1/5}$(We will see later that this is the bandwidth speed that balances variance and squared bias.) Using this bandwidth we can state

$$\begin{aligned}\widehat{r}_n(x) - m(x)\widehat{f}_n(x) &= (\widehat{r}_n(x) - r(x)) - m(x)(\widehat{f}_n(x) - f(x)) \\ &= \mathcal{O}_{\mathbb{P}}(n^{-2/5}) + m(x)\mathcal{O}_{\mathbb{P}}(n^{-2/5}) \\ &= \mathcal{O}_{\mathbb{P}}(n^{-2/5}),\end{aligned}$$

$$\begin{aligned}(\widehat{m}(x) - m(x))(f(x) - \widehat{f}_n(x)) &= o_{\mathbb{P}}(1) + \mathcal{O}_{\mathbb{P}}(n^{-2/5}) \\ &= \mathcal{O}_{\mathbb{P}}(n^{-2/5}).\end{aligned}$$

Hence, in Equation (8) the leading term in the distribution of $\widehat{m}(x) - m(x)$ is $\frac{\widehat{r}_n(x) - m(x)\widehat{f}_n(x)}{f(x)}$, since the second term in (8) is for $(n \to \infty)$ of smaller order in probability. The MSE of this leading term is as follows:

$$\begin{aligned}\frac{1}{f^2(x)}\mathbb{E}[(\widehat{r}_n(x) - m(x)\widehat{f}_n(x))^2] &= \frac{1}{(nhf(x))^2}\mathbb{E}\left[\left(\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)(Y_i - m(x))\right)^2\right] \\ &= \frac{1}{nh_n^2 f^2(x)} Var\left[K\left(\frac{x - X}{h_n}\right)(Y - m(x))\right] \\ &\quad + \frac{1}{h^2 f^2(x)}\mathbb{E}^2\left[K(\frac{x - X}{h_n})(Y - m(x))\right].\end{aligned}$$

After some calculations very similar to those we did for density estimation we arrive at the following approximation: for $h \to 0, nh \to \infty$

$$MSE[\widehat{m}(x)] = \frac{1}{nh_n}\frac{\sigma^2(x)}{f(x)}\|K\|_2^2 + \frac{h_n^4}{4}\left(m''(x) + 2\frac{m'(x)f'(x)}{f(x)}\right)^2 \mu_2^2(K) + o(nh_n^{-1}) + o(h_n^4). \quad (9)$$

Note the following points:

(1) The MSE is order $\mathcal{O}(n^{-4/5})$, when we choose $h_n \sim n^{-1/5}$.

(2) The first summand of (9) describes the asymptotic variance of $\widehat{m}(x)$. This formula for the variance is a function of the marginal density $f(x)$ and the conditional variance $\sigma^2(x)$ and agrees with the intuitive assumption that the regression curve is more stable in those areas where we have plenty of observations.

(3) The second summand corresponds to the squared bias of $\widehat{m}(x)$ and is either dominated by the second derivative $m^{''}(x)$, when we are near to a local extremum of $m(x)$ or by the first derivative $m^{'}(x)$, when we are near to a deflection point of $m(x)$.

4.1. Asymptotic Nomality

The asymptotic distribution of $\widehat{m}(x)$ is stated in the following theorem.

Theorem 4.1.

Suppose that

- $\int |K(u)|^{2+\eta} du < \infty$ for some $\eta > O$.
- $h_n \sim n^{-1/5}$.
- m and f are twice differentiable.
- The distinct points $x_1, \dots, x_k$ are continuity points of $\sigma^2(x)$ and of $\mathbb{E}[|Y|^{2+\eta}|X = x]$ and $f(x_j) > 0$, for $j = 1, 2, \dots, k$.

Then, the Nadaraya-Watson kernel smoother $\widehat{m}(x_j)$ at the k different locations $x_1, \dots, x_k$ converges in distribution to a multivariate normal random vector with mean B and identity covariance matrix,

$$\left\{ \sqrt{nh} \frac{\widehat{m}(x_j) - m(x_j)}{\sqrt{\frac{\sigma^2(x_j)\|K\|_2^2}{f(x_j)}}} \right\}_{j=1}^{k} \rightsquigarrow \mathcal{N}(B, I),$$

where

$$B = \left\{ \mu_2(K)[\frac{m^{''}(x_j) + 2m^{'}(x_j)f^{'}(x_j)}{f(x_j)}] \right\}_{j=1}^{n}.$$

Proof:

The proof is given in Ferraty et al. (2003). ■

Theorem 4.2.

Let the regularity assumptions (1)-(7) on the density and the kernel be satisfied, if $h_n \simeq (\ln n/n)^{1/6}$ then

$$\sup_{x \in \mathbb{R}} |\widehat{r}_n(x) - r(x)| = \mathcal{O}_{\mathbb{P}} \left(\left(\frac{\ln n}{n} \right)^{1/3} \right),$$

and

$$\sup_{x\in\mathbb{R}} |\widehat{r}_n(x) - r(x)| = \mathcal{O}_{a.s}\left(\left(\frac{\ln n}{n}\right)^{1/3}\right).$$

Proof:

The proof is identical to the ones of theorems Theorem 3.1 and Theorem 3.3, but uses propositions Proposition 4.3 and 4.4 below instead of propositions Proposition 3.6 and 3.7.

Proposition 4.3.

Let the regularity assumptions (1)-(7) on the density and the kernel be satisfied, then for a compact set $D \subset (0,1)^2$ and $h_n \simeq (\ln n/n)^{1/6}$, one has

$$\sup_{(u,v)\in D} |\widehat{c}_n(u,v) - c_n(u,v)| = \mathcal{O}_{\mathbb{P}}\left(\left(\frac{\ln n}{n}\right)^{1/3}\right) = \mathcal{O}_{a.s}\left(\left(\frac{\ln n}{n}\right)^{1/3}\right).$$

Proposition 4.4.

Let the regularity assumptions (1)-(7) on the density and the kernel be satisfied, then for a compact set $D \subset (0,1)^2$, $h_n \to 0$ and $nh_n^3/\ln n \to \infty$ entails

$$\sup_{(x,y)\in D} |\widehat{c}_n(F_n(x), G_n(y)) - c_n(F(x), G(y))| = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{nh_n^4} + \frac{\ln n}{\sqrt{n}}\right),$$

$$\sup_{(x,y)\in D} |\widehat{c}_n(F_n(x), G_n(y)) - c_n(F(x), G(y))| = \mathcal{O}_{a.s}\left(\frac{\ln n\sqrt{\ln\ln n}}{\sqrt{n}} + \frac{\ln\ln n}{nh_n^4}\right).$$

■

Remark 4.5.

Since the copula density c has a compact support $[0,1]^2$, our estimator may suffer from bias issues on the boundaries, i.e., in the tails of X and Y . To correct these issues, one could apply one of the several known techniques to reduce the bias of the kernel estimator on the edges (see e.g Fan et al. (1996) Chapter 5.5, boundary kernels, reflection, transformation and local polynomial fitting). In the tail of the distribution of X, this bias issue in the copula density estimator is balanced by the improved variance, as shown below.

(1) **Variance**
The variance of our estimator involves a product of the density $g(y)$ of Y by the conditional density $f(y|x)$,

$$na_n^2 Var(\widehat{f}(y|x) \approx g(y)f(y|x) = g^2(y)c(F(x), G(y)),$$

where $= Var(\widehat{f}(y|x)) = 1/(na_n g(y)f(y|x)\|\|K\|\|_2^2 + o(1/(na_n^2))$, whereas competitors involve the ratio of $f(y|x)$ by the density $f(x)$ of X

$$\frac{f(y|x)}{f(x)} = \frac{g(y)}{f(x)}c(F(x), G(y)).$$

It is a remarkable feature of the estimator we propose, that its variance does not involve directly $f(x)$, as is the case for the competitors, but only its contribution to Y , through the copula density. This reflects the ability announced in the introduction of the copula representation to have effectively separated the randomness pertaining to Y alone, from the dependence structure of (X, Y). Moreover, our estimator also does not suffer from the unstable nature of competitors who, due to their intrinsic ratio structure, get an explosive variance for small value of the density $f(x)$, making conditional estimation difficult, e.g. in the tail of the distribution of X.

(2) **Bandwidth selection**
Performance of nonparametric estimators depends crucially on the bandwidths. For conditional density, bandwidth selection is a more delicate matter than for density estimation due to the multidimensional nature of the problem. Moreover, for ratio-type estimators, the difficulty is increased by the local dependence of the bandwidths h_y on h_x implied by conditioning near x. For the copula estimator, a supplemental issue comes from the fact that the pseudo-data $F(X_i), G(Y_i)$ is not directly accessible. Inspection of the AMISE of the copula-based estimator suggest we can separate the bandwidth choice of h for $\widehat{g}(y)$ from the bandwidth choice of an the copula density estimator $\widehat{c}_n$. A rationale for a data-dependent method is to separately select h on the Y_i data alone (e.g. by cross-validation or plug-in), from the an of the copula density c based on the approximate data $F_n(X_i), G_n(Y_i)$. However, such a bandwidth selection would require deeper analysis and we leave a detailed study of a practical data-dependent method for bandwidth selection of the copula-quantile estimator, together with a global and local comparison of the estimators at their respective optimal bandwidths for further research.

5. Examples

In this part we study ordering properties and bounds for conditional distributions with specific dependence models (copulas). In the first example, we study the results for a specific Clayton-Oakes copula.

First, we give the formal definition of distorted distributions.

Definition 5.1.

We say that F_q is a distorted distribution of a distribution function F if

$$F_q(t) = q(F(t)),$$

for all t, where q is a distortion function, that is, $q: [0, 1] \longrightarrow [0, 1]$ is a continuous increasing function such that $q(0) = 0$ and $q(1) = 1$.

We are going to consider three different conditional distributions. The first one is the distribution of the random variable $(Y|X \leq x)$ which can be written as

$$\mathbb{P}(Y \leq y|X \leq x) = \frac{\mathbb{P}(Y \leq y, X \leq x)}{\mathbb{P}(X \leq x)} = \frac{C(F(x), G(y))}{F(x)} = q_1(G(y)),$$

where $F(x) = \mathbb{P}(X \leq x)$ and $G(y) = \mathbb{P}(Y \leq y)$ are the (marginal) distribution functions of X and Y, respectively, and where C is a copula (i.e., C is a restriction of a continuous distribution function with uniform marginal over the interval $(0, 1)$). The distortion function is given by

$$q_1(u) = \frac{C(F(x), u)}{F(x)}.$$

Example 5.2.

Let us consider a random vector (X, Y) with the following Clayton-Oakes (distributional) copula:

$$C(u, v) = \frac{uv}{u + v - uv},$$

for $0 \leq u,\ v \leq 1$. Then, the distortion function of $(Y|X \leq x)$ is

$$q_1(u) = \frac{u}{F(x) + u - uF(x)}.$$

Even more, as

$$C(F(x), u) = \frac{uF(x)}{F(x) + u - uF(x)},$$

is concave in u in the interval $(0, 1)$.

In the following example we study the results for all the Farlie-Gumbel-Morgenstern (FGM) bidimensional copulas which include positive and negative (weak) dependence.

Example 5.3.

Let us consider the random vector (X, Y) with the following FGM distributional copula:

$$C(u, v) = uv[1 + \theta(1 - u)(1 - v)],$$

for $0 \leq u,\ v \leq 1$ and $-1 \leq \theta \leq 1$. In this case, the survival copula coincides with the distributional copula, that is, $\overline{C}(u, v) = C(u, v)$.

The survival copula $\overline{C}$ is determined by the "distributional" copula C (and vice versa) by the following relationship

$$\overline{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

The function

$$C(F(x), u) = uF(x)[1 + \theta(1 - u)\overline{F}(x)], \text{ where } \overline{F}(x) = 1 - F(x),$$

is concave (convex) in u in the interval $(0, 1)$ for $\theta > 0$ $(\theta < 0)$.

Example 5.4.

If the copula C is a strict Archimedean copula then it can be written as

$$C(u,v) = \phi^{-1}(\phi(u) + \phi(v)),$$

where ϕ is a strict generator (see, e.g., Nelsen [28, p. 112]), that is, ϕ is a strictly decreasing continuous and convex function from $[0,1]$ to $[0,\infty]$ such that $\phi(0) = \infty$ and $\phi(1) = 0$. Then,

$$\phi(C(u,v)) = \phi(u) + \phi(v),$$

and, if we assume that ϕ is differentiable, we have

$$\frac{\partial C(u,v)}{\partial u} = \frac{\phi'(u)}{\phi'(C(u,v))},$$

whenever $\phi'(C(u,v)) < 0$.

Example 5.5.

(1) **On the asymptotic equivalence with Stute's smooth k-Nearest Neighbor regression estimator**

Now, we explore the connection between the conditional mean predictor and Yang and Stute's Yang (1981), Stute (1982) smoothed k-Nearest neighbor estimator of the regression.
Indeed, we recall the different k-Nearest Neighbor estimators of the regression.

- The classical k-NN estimator of the regression has been introduced by Loftsgaarden and Queensbury Loftsgaarden et al. (1965). Fix x, and reorder the sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ according to the increasing values of $|X_i - x|$ as

$$(X_{(1,n)}(x), Y_{(1,n)}(x)), (X_{(2,n)}(x), Y_{(2,n)}(x)), \ldots, (X_{(n,n)}(x), Y_{(n,n)}(x)).$$

Then, the estimator is

$$\widetilde{r}^{NN}(x) = \frac{1}{k_n} \sum_{i=1}^{k_n} Y_{(i,n)}(x).$$

For the density, an estimator is defined analogously as

$$\widehat{f}(x) = \frac{\mu_n[B(x, R(k,x))]}{\lambda[B(x, R(k,x))]},$$

where μ_n stands for the empirical measure, $B(x,\varepsilon)$ the ball of radius ε centered at x and $R(k,x)$ is the distance from x to the k_nth nearest of $X_1, X_2, \ldots, X_n$. Stone (1977) showed its universal consistency. See also Liero (1993). Moore et al. (1977) defined a generalized version as

$$\widehat{f}(x) = \frac{1}{nR(k,x)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{R(k,x)}\right).$$

- Yang (1981) and Stute (1984) smoothed version of the k-NN estimator is defined as

$$r^{NN}(x) = \frac{1}{na_n} \sum_{i=1}^{n} Y_i K\left(\frac{F_n(x) - F_n(X_i)}{a_n}\right).$$

The fact that r^{NN} is a smoothed version of the k-NN estimator can be seen with the kernel $K = \mathbf{1}_{[-1/2,1/2]}$. In this case, $r^{NN}(x_0)$ is the average number of Y_i for which, when $X_i \geq x_0$ (say), there exists no more than $k_n := nh_n/2$ X_j values with $x_0 \leq X_j$ and such that $Xj < X_i$.

- with the preceding discussions in sections Section 2 and Section 3 on the asymptotic deficiency of the empirical c.d.f. with respect to its kernel smoothed version in mind, we can define the doubly smoothed Nearest-Neighbour estimator of the regression function as

$$\widehat{r}^{NN}(x) = \frac{1}{na_n} \sum_{i=1}^{n} Y_i K\left(\frac{\hat{F}_n(x) - \hat{F}_n(X_i)}{a_n}\right),$$

where $\widehat{F}_n$ the kernel smoothed estimator of the c.d.f. F.

(2) **On conditional empirical distribution function**
A more general way to view the resemblances and dissimilarities between the quantile copula estimator and Stute's one, is to have an approach based on conditional empirical process. The conditional empirical process indexed by the function ϕ is defined as

$$\mathbb{G}(y|x) = \sum_{i=1}^{n} w_{ni}(x, X_1, \ldots, X_n)\phi(Y_i, y),$$

where $(w_{ni})_{1\leq i\leq n}$ is a sequence of weights. In particular, for the class of functions $\phi = \mathbf{1}_{Y_i \leq y}$, the conditional distribution function can be written as

$$\widehat{F}(y|x) = \sum_{i=1}^{n} w_{ni}(x, X_1, \ldots, X_n)\mathbf{1}_{Y_i \leq y}.$$

- With the Nadara-Watson weights defined as

$$w_{ni} = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)},$$

the Nadaraya-Watson regression and double kernel conditional density estimators are defined respectively as

$$\widehat{f}^{DK}(y|x) = \int K_h(y - t)d\widehat{F}(t|x)$$

$$\text{and } \widehat{r}^{DK}(y|x) = \int t d\widehat{F}(t/x).$$

- In Stute's nearest neighbour approach, the weights are different,

$$w'_{ni} = \frac{1}{a_n} K\left(\frac{F_n(x) - F_n(X_i)}{a_n}\right),$$

and the estimators are defined analogously, as

$$\widehat{f}(y|x) = \int K_a(y - t)d\hat{F}(t/x)$$

$$\text{and } \widehat{r}(y|x) = \int t d\widehat{F}(t|x).$$

Notice that the Nadaraya-Watson weights sum to 1, whereas the nearest neighbour weights asymptotically sum up to 1.

- In the quantile copula approach, the weights are also w'_{ni} as Stute's approach, but the density and regression estimators are defined in a slightly different manner, as

$$\widehat{f}(y|x) = \int \widehat{g}(y) K_a(G_n(y) - G_n(t)) d\widehat{F}(t|x)$$

$$\text{and } \widehat{r}(y|x) = \int y \widehat{g}(y) d\widehat{F}(y|x).$$

(3) **Local Polynomial estimators**

Another approach would be to estimate the copula density by the Local polynomial method of Fan et al. (1996). Its extension in the multivariate case has been investigated by Abdous et al. (2001). See also Abdous et al. (2005). Recall that the local polynomial estimator of the copula function is defined as the minimiser of the following score,

$$L_n(x) = \int_{[0,1]^2} K_H(u - x)[C_n(u) - P(u - x)]^2 du,$$

where $C_n(\cdot) = n^{-1} \sum_{i=1}^n \mathbf{1}_{(F_n(X_i), G_n(Y_i)) \leq \cdot}$ is the empirical cumulative distribution function based on the approximate observations, $K_H(u) = |H|^{-1} K(H^{-1}u)$ a bivariate kernel with bandwidth matrix H, P is a multivariate polynomial of given degree, and x and u are bivariate in that context. Note that the polynomial has to be of degree at least one to estimate the copula density and that one can replace the empirical c.d.f C_n by any smooth estimate converging a.s. to it.

An alternative approach would be to replace C_n by the empirical measure $\mu_n(\cdot, \cdot) = n^{-1} \sum_{i=1}^n \delta_{F_n(X_i)}(\cdot) \delta_{G_n(Y_i)}(\cdot)$ to estimate directly the copula density and its derivatives.

The local polynomial method, when the integral to be minimised is restricted to the support of the function to be estimated, has the advantage of being free of boundary bias, thus correcting the bias issue of the kernel method.

Estimation of the conditional cumulative distribution function

The Quantile transform and copula representation approach which lead to the proposed estimator of the conditional density can also be used to estimate the conditional cumulative distribution function $F(y|x)$, we have two possible approaches:

(1) **Approach by integration of the conditional density**

From the estimator of the conditional density, one can integrate to obtain an estimate of the conditional c.d.f.:

$$F_{Y|X}(x, y) = \int_{-\infty}^{y} f_{Y|X}(x, u) du \text{ entails } \widehat{F}_{Y|X}(x, y) \int_{-\infty}^{y} \widehat{f}_{Y|X}(x, u) du,$$

and the latter integral can be computed by numerical integration. One can also get an approximate explicit formula by noting that,

$$F_{Y|X}(x, y) = \int_{-\infty}^{y} g(t) c(F(x), G(t)) dt = \mathbb{E}\Big[\mathbf{1}_{Y \leq y} c(F(x), G(Y))\Big],$$

and replacing the expectation by the empirical mean and the unknown c, F, G by their respective estimators $\widehat{c}_n$, $\widehat{F}_n$, $\widehat{G}_n$:

$$\widehat{F}_{Y|X}(x,y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \le y}\widehat{c}_n(F_n(x), G_n(Y_i)).$$

(2) **Direct approach**
From Sklar's copula formula, one has by a change of variable that

$$F_{Y|X=x}(x,y) = \int_{-\infty}^{y} g(t)c(F(x), G(t))dt = \int_{0}^{G(y)} c(F(x), v)dv.$$

To construct an estimator, one may replace the unknown quantities c, F, G by the respective estimators $\widehat{c}_n$, F_n, G_n to obtain

$$\widehat{F}_{Y|X=x}(x,y) = \int_{0}^{G_n(y)} \widehat{c}_n(F_n(x), v)dv = \frac{1}{na_n^2}\sum_{i=1}^{n} K\left(\frac{F_n(x) - F_n(X_i)}{a_n}\right)\int_{0}^{G_n(y)} K\left(\frac{v - G_n(Y_i)}{a_n}\right)dv.$$

By setting $K^1(t) = \int_{-\infty}^{t} K(v)dv$, the latter estimator writes, after integration, as

$$\widehat{F}_{Y|X=x}(x,y) = \frac{1}{na_n}\sum_{i=1}^{n} K\left(\frac{F_n(x) - F_n(X_i)}{a_n}\right)\left[K^1\left(\frac{G_n(y) - G_n(Y_i)}{a_n}\right) - K^1\left(\frac{-G_n(Y_i)}{a_n}\right)\right].$$

5.1. Finite sample numerical simulation

Although the proposed estimator seems to compare favorably asymptotically, some pitfalls linked to the copula density estimation may show up in the practical implementation

5.2. Model and comparison results

Many copula densities exhibit infinite values at their corners. Therefore, to avoid that $(F_n(X_i), G_n(Y_i))$ be equal to $(1,1)$, we change the empirical distribution functions F_n and G_n to $n/(n+1)F_n$ and $n/(n+1)G_n$ respectively.

We simulated a sample of $n = 100$ variables (X_i, Y_i), from the following model: X, Y is marginally distributed as $\mathcal{N}(0,1)$ and linked via Frank Copula.

$$C(u,v,\theta) = \frac{\ln[(\theta + \theta^{u+v} - \theta^u - \theta^v)/(\theta - 1)]}{\ln\theta},$$

with parameter $\theta = 100$.

We restricted ourselves to simple, fixed for all x, y, rule-of-thumb methods based on Normal reference rule to get a first picture. For the selection of an of the copula density estimator, we applied

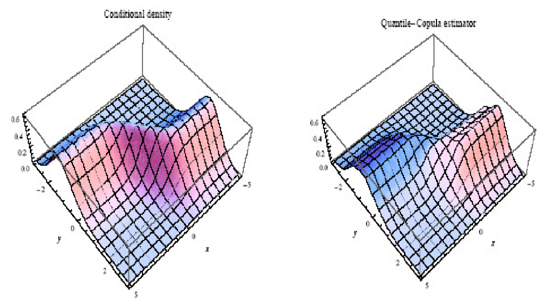

Figure1. 3D Plots. From left to right, top to bottom: true density, quantile-copula estimator

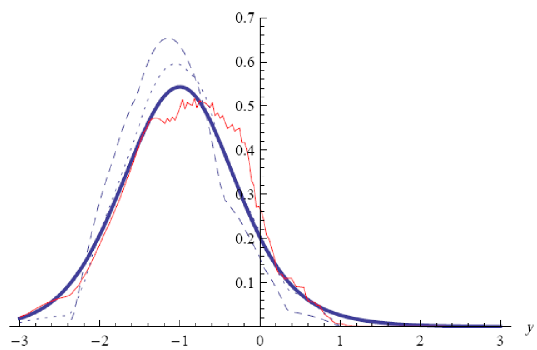

Figure 2. Comparison at $x=2$: conditional density = thick curve, Quantile copula=continuous line, double kernel=dotted curve, local polynomial=dashed curve

Scott's Rule on the data $F_n(X_i)$. We used Epanechnikov kernels for $\widehat{g}(y)$ and the other estimators. We plotted the conditional density along with its estimations on the domain $x \in [-5, 5]$ and $y \in [-3, 3]$ on figure 1. A comparison plot at $x = 2$ is shown on figure 2.

6. Conclusion

In this paper we established the asymptotic properties of a regression model via copula function approach, and to study the asymptotic normality of such model, it will be interesting in further work to study the recursive estimation, it is also important to study the asymptotic properties of a conditional copula model.

In this conclusive section, we would like to sketch some perspectives for further research and possible applications of the proposed estimator. These are developed in a more or less lengthy manner, depending on the current degree of advancement of our ongoing research.

The proposed estimator could be refined by using some more sophisticated methods of estimation of the copula density.

At last, possible fields of application such as extremes or missing data where we believe the proposed estimator could be an interesting starting.

In the mixing framework, analogues of the Chung-Smirnov property and convergence results of the kernel density estimator do exist. By coupling arguments as in Rio (2000), one should be able to extend the estimator in the dependent framework. A case of particular interest is when the X, Y variables corresponds to the X_n, X_{n+1} of a stationary Markov chain, which gives an estimate of the transition density. Such an estimate should serve as a building block to make inference, tests and prediction in fields such as e.g. econometrics.

The method of estimation would be to combine results and method of standard extreme theory together with this nonparametric approach, and could be briefly sketched as follows:

Such a study could be of practical importance in the following fields.

(1) Environmental applications: For preventing floods, one can be interested in under- standing the impact of big waves occurring in windstorms on the water level, or the impact of the amount of rain on the flow of a river. See e.g. Haan et al. (1999).
Other possible applications could be energy production given wind strength for windmills, electrical consumption of households given temperature, impact of a given factor on pollution, etc.
(2) Insurance and finance applications: One can imagine this approach could be useful in finance, e.g. to detect the possibility of bankruptcy of an insurance company from a large claim, occurring with small frequency.
(3) Reliability applications: to assess how the failure of one component would impact the failure of another one, etc.

Acknowledgment

The authors would like to express their gratitude to the referees and chief editor for their valuable suggestions which have considerably improved the earlier version of the paper.

REFERENCES

Abdous, B. and Bensaid, E. (2001). Multivariate local polynomial fitting for a probability distribution function and its partial derivatives, J. Nonparametr. Statist., Vol. 13, No. 1, pp. 77–94.

Abdous, B. and Ghoudi, K. (2005). Non-parametric estimators of multivariate extreme dependence functions, J. Nonparametr. Stat., Vol. 17, No. 8, pp. 915–935.

Biau, G. and Wegkamp, M.H. (2006). A note on minimum distance estimation of copulas densities, Statist. Probab. Lett., Vol. 73, pp. 105–114.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes. Estimation and prediction*, second ed., **110** of Lecture Notes in Statistics. Springer-Verlag, New York.

Carmona, R.A. (2004). *Statistical Analysis of Financial Data in S-PLUS*, Springer-Verlag, New York.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, John Wiley and Sons, Chichester.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance, Acad. Roy. Belg. Bull. Cl. Sci., Vol. 65, No. 5-6, pp. 274–292.

Deheuvels, P. and Kolmogorov-Smirnov, A. (1981). type test for independence and multivariate samples, Rev. Roumaine Math. Pures Appl., Vol. 26, No. 2, pp. 213–226.

Embrechts, P., Lindskog, F. and McNeil, A. (2003). *Modelling dependence with copulas and applications to risk management*, In: Handbook of Heavey Tailed Distributions in Finance, S.T. Rachev (ed.). Elsevier, Amsterdam.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, Vol. 66 of Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, Biometrika, Vol. 83, No. 1, pp. 189–206.

Faugeras, O.P. (2008). *Some contributions to statistical prediction*, Theses, Université Pierre et Marie Curie - Paris VI.

Fermanian, J.D. (2005). Goodness-of-fit tests for copulas, J. Multivariate Anal., Vol. 95, No. 1, pp. 119–152.

Fermanian, J.D., Radulovïc, C.D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes, Bernoulli, Vol. 10, No. 5, pp. 847–860.

Fermanian, J.D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series, Journal of Risk, Vol. 5, No. 4, pp. 25–54.

Fermanian, J.D. and Scaillet, O. (2005). *Some statistical pitfalls in copula modelling for financial applications*, In: E. Klein (Ed.), Capital formation, Gouvernance and Banking. Nova Science Publishing, New York.

Ferraty, F. and Vieu, P. *Statistique fonctionnelle: modèles de régression pour variables uni, multi et infiniment dimensionnées*, *Publications du LSP*, Toulouse, France, 2001-03.

Fisher, N.I. (1997). *Copulas*, In: Encyclopedia of Statistical Sciences, Updated Vol. 1, S. Kotz, C. B. Read, D. L. Banks (eds.), pp. 159–164. John Wiley and Sons, New York.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function, Comm. Statist. Theory and Methods, Vol. 19, No. 2, pp. 445–464.

De Haan, L. and Sinha, A.K. (1999). Estimating the probability of a rare even,. Ann. Statist., Vol. 27, No. 2, pp. 732–759.

Hutchinson, T.P. and Lai, C.D. (1990). *Continuous Bivariate Distributions: Emphasising Applications*, Rumsby Scientific Publishing, Adelaide.

The International Actuarial Association. (2004). *A global framework for insurer solvency assessment*, Research Report. www.actuaries.org.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

Liero, H. (1993). A note on the asymptotic behaviour of the distance of the knth nearest neighbour, Statistics, Vol. 24, No. 3, pp. 235–243.

Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function, Ann. Math. Statist., Vol. 36, pp. 1049–1051.

Mikosch, T. (2006). Copulas: Taies and facts, Extrêmes, Vol. 9, No. 18, pp. 3–22.

Moore, D.S. and Yackel, J.W. (1977). Consistency properties of nearest neighbor density function estimators, Ann. Statist., Vol. 5, No. 1, pp. 143–154.

Nelsen, R.B. (1999). *An Introduction to Copulas*, Springer-Verlag, New York.

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

Parzen, E. (1962). On estimation of a probability density function and mode, Ann. Math. Statist., Vol. 33, pp. 1065–1076.

Pickands, J. (1975). Statistical inference using extreme order statistics, Ann. Statist., Vol. 3, pp. 119–131.

Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*, Vol. 31 of Mathématiques & Applications (Berlin) (Mathematics & Applications). Springer-Verlag, Berlin.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics, Ann. Statist., Vol. 4, No. 5, pp. 912–923.

Sancetta, A. (2003). *Nonparametric estimation of multivariate distributions with given marginals: L_2 theory*, Cambridge Working papers in Economics No. 0320.

Sancetta, A. and Satchell, S. (2004). The Barnestien coplua and its application to modelling and approximation of multivariate distributions, Econometric theory, Vol. 20, pp. 535–562.

Scott, D.W. (1992). *Multivariate density estimation. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*, John Wiley & Sons Inc., Theory, practice, and visualization, A Wiley-Interscience Publication, New York.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8, pp. 229–231.

Stone, C.J. (1977). Consistent nonparametric regression, Ann. Statist., vol. 5, No. 4, pp. 595–645.

Stute, W. (1982). The oscillation behavior of empirical processes, Ann. Probab., Vol. 10, No. 1, pp. 86–107.

Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates, Ann. Statist., Vol. 12, No. 3, pp. 917–926.

Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes with applications to statistics*, Springer Series in Statistics. Springer-Verlag, New York.

Yang, S.S. (1981). Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function, J. Amer. Statist. Assoc., Vol. 76, No. 375, pp. 658–662.