Probit transformation for nonparametric kernel estimation of the copula density

GERY GEENENS[1], ARTHUR CHARPENTIER[2] and DAVY PAINDAVEINE[3]

[1]*School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia.*
E-mail: ggeenens@unsw.edu.au
[2]*Département de mathématiques, Université du Québec à Montréal, Canada*
[3]*ECARES, Université Libre de Bruxelles, Belgium*

Copula modeling has become ubiquitous in modern statistics. Here, the problem of nonparametrically estimating a copula density is addressed. Arguably the most popular nonparametric density estimator, the kernel estimator is not suitable for the unit-square-supported copula densities, mainly because it is heavily affected by boundary bias issues. In addition, most common copulas admit unbounded densities, and kernel methods are not consistent in that case. In this paper, a kernel-type copula density estimator is proposed. It is based on the idea of transforming the uniform marginals of the copula density into normal distributions via the probit function, estimating the density in the transformed domain, which can be accomplished without boundary problems, and obtaining an estimate of the copula density through back-transformation. Although natural, a raw application of this procedure was, however, seen not to perform very well in the earlier literature. Here, it is shown that, if combined with local likelihood density estimation methods, the idea yields very good and easy to implement estimators, fixing boundary issues in a natural way and able to cope with unbounded copula densities. The asymptotic properties of the suggested estimators are derived, and a practical way of selecting the crucially important smoothing parameters is devised. Finally, extensive simulation studies and a real data analysis evidence their excellent performance compared to their main competitors.

Keywords: boundary bias; copula density; local likelihood density estimation; transformation kernel density estimator; unbounded density

1. Introduction

For the last two decades copula modeling has emerged as a major research area of statistics. A bivariate copula function C is the joint cumulative distribution function (c.d.f.) of a bivariate random vector whose marginals are Uniform over [0, 1], that is, $C : \mathcal{I} \doteq [0, 1]^2 \to [0, 1] :$ $(u, v) \to C(u, v) = \mathbb{P}(U \leq u, V \leq v)$, where $U \sim \mathcal{U}_{[0,1]}$, $V \sim \mathcal{U}_{[0,1]}$. Copulas arise naturally as a mere consequence of two well-known facts. First, the *probability-integral transform* result: if $X \sim F_X$ is continuous, then $F_X(X) \sim \mathcal{U}_{[0,1]}$; and second, *Sklar's theorem* [49]: for any continuous bivariate c.d.f. F_{XY}, there exists a unique function C such that

$$F_{XY}(x, y) = C\big(F_X(x), F_Y(y)\big) \qquad \forall (x, y) \in \mathbb{R}^2, \tag{1.1}$$

where F_X and F_Y are the marginals of F_{XY}. From the above, this function C is, indeed, a copula, called the copula of F_{XY}. It describes how F_X and F_Y 'interact' to produce the joint F_{XY} and

clearly disjoints the marginal behaviors of X and Y from their dependence structure. See Joe [33] and Nelsen [42] for book length treatment of the foregoing ideas. More compact reviews include [30] and [16]. Today, copulas are used extensively in statistical modeling in all areas, including quantitative finance and insurance. Therefore, empirically estimating a copula function from a bivariate sample $\{(X_i, Y_i)\}_{i=1}^n$ drawn from F_{XY} has become an important problem of modern statistical modeling.

Of course, estimating C essentially amounts to fitting a bivariate distribution, for what many parametric families have been suggested and studied: Gaussian, Student-t, Clayton, Frank or Gumbel copulas among others (see again [33] or [42] for details). These parametric models have formed the main body of the literature in the field so far. However, they suffer from the usual lack of flexibility of parametric approaches and the induced risk of misspecification. For instance, it has been argued that the main reason behind the 2009 global financial crisis was a reckless usage of the Gaussian copula [45]. There is, therefore, a tremendous need for flexible nonparametric copula models, making no rigid assumptions on the underlying distributions. An early step in that direction was the empirical copula devised by Deheuvels [12]. The related empirical copula process was studied further in [19,47,50] and [6], and turns out to be the cornerstone of a variety of nonparametric copula-based procedures, see e.g. [25,26,29] or [37], to cite only a few. Moreover, Fermanian and Scaillet [20], Chen and Huang [10] and Omelka, Gijbels and Veraverbeke [43] studied kernel methods to obtain flexible smooth estimates of the bivariate c.d.f. C.

Usually, though, a distribution is more readily interpretable in terms of its density function than directly through its c.d.f., and a copula is no different. Assume that the bivariate c.d.f. C is absolutely continuous. Then, its associated *copula density* is

$$c(u, v) = \frac{\partial^2 C}{\partial u \, \partial v}(u, v)$$

for $(u, v) \in \mathcal{I}$. This paper addresses the problem of nonparametrically estimating c, for what kernel methods again appear natural. This approach was pioneered in [2] and [28], and arguably remains very attractive compared to its competitors, such as splines [35,48], wavelets [1,24], Bernstein polynomials [4,5,32] or others [44], for its simplicity.

Three factors make kernel estimation of c not standard, though. First, a major concern is that kernel estimators suffer from boundary bias problems. Given the bivariate sample $\{(U_i = F_X(X_i), V_i = F_Y(Y_i))\}_{i=1}^n$, the standard kernel estimator for c, say $\hat{c}^*$, at $(u, v) \in \mathcal{I}$ would be [51], Chapter 4

$$\hat{c}^*(u, v) = \frac{1}{n|\mathbf{H}_{UV}|^{1/2}} \sum_{i=1}^n \mathbf{K}\left(\mathbf{H}_{UV}^{-1/2}\begin{pmatrix} u - U_i \\ v - V_i \end{pmatrix}\right), \tag{1.2}$$

where $\mathbf{K} : \mathbb{R}^2 \to \mathbb{R}$ is a kernel function and $\mathbf{H}_{UV}$ is a bandwidth matrix. However, estimator (1.2) is known not to be consistent along boundaries of $\mathcal{I}$: standard arguments show that $\mathbb{E}(\hat{c}^*(u, v)) = \frac{1}{4}c(u, v) + O(h)$ at corners and $\mathbb{E}(\hat{c}^*(u, v)) = \frac{1}{2}c(u, v) + O(h)$ on the borders. In fact, $\hat{c}^*$ does not 'feel' the boundaries and places through $\mathbf{K}$ positive mass beyond them. Although some papers ignored these issues [18,20,46], it is clear that accurate estimation of c calls for some boundary correction. Such corrections have indeed been proposed, e.g. mirror reflection [28] or the usage of boundary kernels [10], but with mixed results.

Secondly, kernel estimators are not consistent for unbounded densities. Yet, many copula densities of interest are unbounded: even in the apparently easy case of a bivariate Normal vector with moderate correlation, the copula density is unbounded in two of the corners of $\mathcal{I}$. Therefore, any good estimator of c should be able to cope with such unboundedness. Finally, estimating c cannot be made from a genuine random sample from its c.d.f. C, as C is the distribution of $(U, V) = (F_X(X), F_Y(Y))$ and F_X and F_Y are typically unknown. Hence, the observations (U_i, V_i) are unavailable, and estimator (1.2) is, in fact, infeasible. Let $\hat{F}_{Xn}(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j \leq x\}}$ be the empirical c.d.f. of X, and similarly for $\hat{F}_{Yn}$. Define

$$\hat{U}_i = \frac{n}{n+1} \hat{F}_{Xn}(X_i) \quad \text{and} \quad \hat{V}_i = \frac{n}{n+1} \hat{F}_{Yn}(Y_i) \tag{1.3}$$

the 'pseudo-observations' (the rescaling by $n/(n+1)$, aiming at keeping $\hat{U}_i$ and $\hat{V}_i$ in the interior of $[0, 1]$, is customary). Then, it is common practice to treat the pseudo-sample $\{(\hat{U}_i, \hat{V}_i)\}_{i=1}^{n}$ as a sample from C and to use it instead of the 'true' sample $\{(U_i, V_i)\}_{i=1}^{n}$, although this may affect the statistical properties of the ensuing estimators [8,27].

This paper proposes and studies a new, kernel-type estimator of the copula density c. It is based on the idea recently suggested in [22] for estimating univariate densities supported on the unit interval. Based on the transformation method, it takes the constrained nature of the support into account from the outset, that is, without relying on ad hoc boundary corrections (reflection, boundary kernels, etc.). In short, the initial $[0, 1]$-supported variables U and V are transformed through the probit function into variables whose supports are unconstrained, the density in the transformed domain is estimated free from boundary issues via local likelihood methods and an estimate of the initial density on $[0, 1]^2$ is obtained by back-transformation. This method appears very natural and, in Geenens's [22] univariate case, lead to estimators superior to their main competitors in the simulation studies for a wide range of density shapes, including for unbounded densities. The idea seems, therefore, suitable for estimating copula densities as well.

This paper does not just study a bivariate extension of Geenens's [22] idea to the copula setting, though. The novel contribution is triple. First, it is explained why the probit transformation is the most appropriate in the copula setting, whereas it was just a convenient choice in the univariate framework. Second, when deriving the asymptotic properties of the estimators, the effect of resorting to pseudo-observations is carefully analyzed through Segers's [47] recent developments on the empirical copula process (of course, the effect of pseudo-observations was not relevant to Geenens's [22] estimators). One of the main conclusion of the paper is that using the pseudo-observations (1.3) instead of genuine observations does not affect the statistical properties of the kernel copula density estimators. This holds true under general bandwidth conditions, including the optimal order, unlike previous papers in the field which required sub-optimal bandwidths (see, e.g., Fermanian's [18] Assumption (B0)) to annihilate the effect of using (1.3). Third, an innovative and effective data-driven bandwidth selection procedure is derived for bivariate density estimation. It does not assume a constrained form (e.g., diagonal) for the bandwidth matrix, and the idea can be used as-is for any multivariate kernel density estimation (it is not specific to copula modeling).

Exploring the probit transformation idea in the context of copula density estimation is the topic of Section 2, and several versions of the estimator will be suggested. Their asymptotic

properties will be derived in Section 3, and the above mentioned bandwidth matrix selection rule will be detailed in Section 4. Simulation studies evidencing the very good practical behavior of the probit-transformation estimators (Section 5), a real data analysis (Section 6) and some final remarks (Section 7) conclude the paper.

2. Probit transformation kernel copula density estimation

2.1. Probit transformation

Direct kernel estimation of c is made difficult mainly by the constrained nature of its support $\mathcal{I} = [0, 1]^2$. Now, define $S = \Phi^{-1}(U)$ and $T = \Phi^{-1}(V)$, where Φ is the standard normal c.d.f. and Φ^{-1} is its quantile (probit) function. Given that both U and V are $\mathcal{U}_{[0,1]}$, S and T are both standard normal, which, of course, does not make the vector (S, T) bivariate normal. That will only be the case if the copula of F_{ST}, that is, C itself (copulas are invariant to increasing transformations of their margins [42], Theorem 2.4.3), is the Gaussian copula. The idea is that, if $c(u, v) > 0$ Lebesgue-a.e. over $\mathcal{I}$, (S, T) has unconstrained support $\mathbb{R}^2$ and estimating its density, say f_{ST}, cannot suffer from boundary issues. In addition, due to its normal margins, one can expect f_{ST} to be very smooth and well-behaved, and its estimation easy. In particular, under mild assumptions, f_{ST} and its partial derivatives up to the second order will be seen to be uniformly bounded on $\mathbb{R}^2$, even in the case of unbounded copula density c (Lemma A.1 in the Appendix [23]).

As the copula of F_{ST} is C, $S \sim \mathcal{N}(0, 1)$ and $T \sim \mathcal{N}(0, 1)$, one has, for all $(s, t) \in \mathbb{R}^2$, $F_{ST}(s, t) = \mathbb{P}(S \leq s, T \leq t) = C(\Phi(s), \Phi(t))$ ((1.1) for (S, T)). Differentiating with respect to s and t, one gets

$$f_{ST}(s, t) = c\big(\Phi(s), \Phi(t)\big)\phi(s)\phi(t), \tag{2.1}$$

where ϕ is the standard normal density. Inverting this expression, one obtains, for any $(u, v) \in (0, 1)^2$,

$$c(u, v) = \frac{f_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{2.2}$$

So, any estimator $\hat{f}_{ST}$ of f_{ST} on $\mathbb{R}^2$ automatically produces an estimator of the copula density on int$(\mathcal{I})$:

$$\hat{c}^{(\tau)}(u, v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \tag{2.3}$$

where the superscript (τ) refers to the idea of transformation. When necessary, $\hat{c}^{(\tau)}$ can also be defined at the boundaries of $\mathcal{I}$ by continuity. This transformation-based estimator enjoys many nice properties. Clearly, $\hat{c}^{(\tau)}$ cannot allocate any probability outside $\mathcal{I}$, since $(\Phi^{-1}(u), \Phi^{-1}(v))$ is not defined for $(u, v) \notin \mathcal{I}$. Also, if $\hat{f}_{ST}$ is a *bona fide* density function, in the sense that $\hat{f}_{ST}(s, t) \geq 0$ for all (s, t) and $\iint_{\mathbb{R}^2} \hat{f}_{ST}(s, t)\, ds\, dt = 1$, then automatically, through the changes of variable $u = \Phi(s)$ and $v = \Phi(t)$, $\hat{c}^{(\tau)}(u, v) \geq 0$ for all $(u, v) \in \mathcal{I}$ and

$\iint_{\mathcal{I}} \hat{c}^{(\tau)}(u, v) \, du \, dv = 1$. Finally, if $\hat{f}_{ST}$ is a uniformly (weak or strong) consistent estimator for f_{ST}, that is, $\sup_{(s,t) \in \mathbb{R}^2} |\hat{f}_{ST}(s, t) - f_{ST}(s, t)| \overset{P/\text{a.s.}}{\to} 0$ as $n \to \infty$, $\hat{c}^{(\tau)}$ inherits that uniform consistency on any compact proper subset of $\mathcal{I}$. As f_{ST} is uniformly continuous on $\mathbb{R}^2$ (Lemma A.1), the estimators suggested for it in the next subsections are all uniformly consistent on $\mathbb{R}^2$. Hence, all the estimators for c suggested in this paper are uniformly consistent on any compact proper subset of $\mathcal{I}$.

2.2. The naive estimator

A first natural idea would be to use, as $\hat{f}_{ST}$ in (2.3), the standard kernel density estimator:

$$\hat{f}_{ST}^{*}(s, t) = \frac{1}{n|\mathbf{H}_{ST}|^{1/2}} \sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}_{ST}^{-1/2}\begin{pmatrix} s - S_i \\ t - T_i \end{pmatrix}\right), \tag{2.4}$$

where $\mathbf{K}$ is a kernel function and $\mathbf{H}_{ST}$ is a bandwidth matrix, and $\{(S_i = \Phi^{-1}(U_i), T_i = \Phi^{-1}(V_i))\}_{i=1}^{n}$ is the transformed sample. However, as the (U_i, V_i)'s are unavailable, so are the (S_i, T_i)'s, and one has to use

$$\left\{\left(\hat{S}_i = \Phi^{-1}(\hat{U}_i), \hat{T}_i = \Phi^{-1}(\hat{V}_i)\right)\right\}_{i=1}^{n}, \tag{2.5}$$

the pseudo-transformed sample, instead. The feasible version of (2.4) is, therefore,

$$\hat{f}_{ST}(s, t) = \frac{1}{n|\mathbf{H}_{ST}|^{1/2}} \sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}_{ST}^{-1/2}\begin{pmatrix} s - \hat{S}_i \\ t - \hat{T}_i \end{pmatrix}\right). \tag{2.6}$$

Through (2.3), this directly leads to the following probit transformation kernel copula density estimator:

$$\hat{c}^{(\tau)}(u, v) = \frac{1}{n|\mathbf{H}_{ST}|^{1/2}\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}_{ST}^{-1/2}\begin{pmatrix} \Phi^{-1}(u) - \Phi^{-1}(\hat{U}_i) \\ \Phi^{-1}(v) - \Phi^{-1}(\hat{V}_i) \end{pmatrix}\right). \tag{2.7}$$

This is essentially the estimator suggested in [8], also used as-is in [40], although it was not studied in any details in those two papers. Omelka, Gijbels and Veraverbeke [43] derived the theoretical properties of an estimator for the copula C (not its density) based on the same transformation.

This idea was, however, called 'naive' in [22] in the univariate case, and the same qualifier could be used for (2.7): although designed to fix boundary issues, it does not provide good results close to the borders, as will be seen later. Geenens [22] explained the reasons for that failure, and suggested some remedies. In particular, estimating the density in the transformed domain via local likelihood offers a promising alternative while keeping the intuitive appeal of the probit-transformation estimator.

2.3. Improved probit-transformation copula density estimators

Loader [39] and Hjort and Jones [31] proposed two similar yet different formulations of the local likelihood density estimator. Loader [39] locally approximates the logarithm of the density by a polynomial, whereas Hjort and Jones [31] consider local parametric density modeling. This paper will only make use of Loader's [39] idea, mainly because the asymptotic theory is more transparent. In any case, both formulations share the same advantages and typically yield very similar estimates.

In this setting of estimating f_{ST} from the pseudo-sample $\{(\hat{S}_i, \hat{T}_i)\}_{i=1}^n$, Loader's [39] local likelihood estimator is defined as follows. Around $(s, t) \in \mathbb{R}^2$, $\log f_{ST}$ is assumed to be well approximated by a polynomial of order $p = 1$ (local log-linear) or $p = 2$ (local log-quadratic). Specifically, it is assumed that

$$\log f_{ST}(\check{s}, \check{t}) \simeq a_{1,0}(s, t) + a_{1,1}(s, t)(\check{s} - s) + a_{1,2}(s, t)(\check{t} - t)$$
$$\doteq P_{\mathbf{a}_1}(\check{s} - s, \check{t} - t) \tag{2.8}$$

for $(\check{s}, \check{t})$ 'close' to (s, t) in the first case ($p = 1$), while in the second case ($p = 2$)

$$\log f_{ST}(\check{s}, \check{t}) \simeq a_{2,0}(s, t) + a_{2,1}(s, t)(\check{s} - s) + a_{2,2}(s, t)(\check{t} - t)$$
$$+ a_{2,3}(s, t)(\check{s} - s)^2 + a_{2,4}(s, t)(\check{t} - t)^2 + a_{2,5}(s, t)(\check{s} - s)(\check{t} - t) \tag{2.9}$$
$$\doteq P_{\mathbf{a}_2}(\check{s} - s, \check{t} - t).$$

The vectors $\mathbf{a}_1(s, t) = (a_{1,0}(s, t), a_{1,1}(s, t), a_{1,2}(s, t))$ and $\mathbf{a}_2(s, t) \doteq (a_{2,0}(s, t), \ldots, a_{2,5}(s, t))$ are then estimated by solving a weighted maximum likelihood problem. For either $p = 1, 2$,

$$\tilde{\mathbf{a}}_p(s, t) = \arg\max_{\mathbf{a}_p} \left\{ \sum_{i=1}^n \mathbf{K}\left(\mathbf{H}_{ST}^{-1/2}\begin{pmatrix} s - \hat{S}_i \\ t - \hat{T}_i \end{pmatrix}\right) P_{\mathbf{a}_p}(\hat{S}_i - s, \hat{T}_i - t) \right.$$
$$\left. - n \iint_{\mathbb{R}^2} \mathbf{K}\left(\mathbf{H}_{ST}^{-1/2}\begin{pmatrix} s - \check{s} \\ t - \check{t} \end{pmatrix}\right) \exp\left(P_{\mathbf{a}_p}(\check{s} - s, \check{t} - t)\right) d\check{s}\, d\check{t} \right\}, \tag{2.10}$$

where, as previously, $\mathbf{K}$ is a kernel function and $\mathbf{H}_{ST}$ is a bandwidth matrix. The estimate of f_{ST} at (s, t) is then $\tilde{f}_{ST}^{(p)}(s, t) = \exp(\tilde{a}_{p,0}(s, t))$, for $p = 1, 2$. 'Improved' probit-transformation kernel copula density estimators for $c(u, v)$ follow from (2.3):

$$\tilde{c}^{(\tau, p)}(u, v) = \frac{\tilde{f}_{ST}^{(p)}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \tag{2.11}$$

for $p = 1$ and $p = 2$. The motivation and the advantages of estimating f_{ST} by local likelihood methods instead of raw kernel density estimation are related to the detailed discussion in [22]. Essentially, the boundary behavior of the estimator of c on $\mathcal{I}$ is dictated by the tail behavior of the estimator of f_{ST} on $\mathbb{R}^2$. But, raw kernel estimators are known to work poorly in the tails of densities, with frequent occurrences of 'spurious bumps'. These fluctuations being greatly magnified

by the back-transformation (2.3), the so-produced estimator of c shows a very erratic behavior at the boundaries. On the contrary, local likelihood estimators have favorable tail behavior, with very smooth tail estimates. This translates into accurate and well-behaved estimates for c close to the boundaries when back into the initial domain $\mathcal{I}$. These observations will be confirmed in the next section, where the asymptotic properties of the estimators ('naive' and 'improved') are derived.

3. Asymptotic properties

For simplicity, it will be assumed that $\mathbf{K}$ is a product Gaussian kernel, that is, $\mathbf{K}(z_1, z_2) = \phi(z_1)\phi(z_2)$, and $\mathbf{H}_{ST} = h^2\mathbf{I}$ for some $h > 0$. Note that, in practice, there are reasons to keep an unconstrained, non-diagonal bandwidth matrix $\mathbf{H}_{ST}$. In particular, the copula density is typically stretched along one of the diagonals of $\mathcal{I}$ when X and Y are dependent, which provides a density f_{ST} likewise stretched along one of the 45 degrees lines in $\mathbb{R}^2$. Hence, using a bandwidth matrix directing smoothing in that particular direction is sensible [15], as discussed further in Section 4. That said, theoretical results for that general case would be less tractable than, while qualitatively equivalent to, the simpler case presented below. Note that for that particular kernel $\mathbf{K}$, $\iint \mathbf{K}^2(z_1, z_2)\,dz_1\,dz_2 = (4\pi)^{-1}$ and $\iint z_k^2\mathbf{K}(z_1, z_2)\,dz_1\,dz_2 = 1$, $k = 1, 2$. These quantities frequently arise in the properties of kernel estimators, and direct use of these particular numerical values will be made in the results below. All proofs are to be found in the Appendix [23].

3.1. The naive estimator and an amended version

Consider the naive estimator (2.7) which, with the above specifications of $\mathbf{K}$ and $\mathbf{H}_{ST}$, reduces to

$$\hat{c}^{(\tau)}(u, v) = \frac{1}{nh^2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\ \times \sum_{i=1}^{n} \phi\left(\frac{\Phi^{-1}(u) - \Phi^{-1}(\hat{U}_i)}{h}\right)\phi\left(\frac{\Phi^{-1}(v) - \Phi^{-1}(\hat{V}_i)}{h}\right). \tag{3.1}$$

Given (2.3), it is clear that its statistical properties will entirely depend on those of (2.6), here

$$\hat{f}_{ST}(s, t) = \frac{1}{nh^2}\sum_{i=1}^{n}\phi\left(\frac{s - \hat{S}_i}{h}\right)\phi\left(\frac{t - \hat{T}_i}{h}\right). \tag{3.2}$$

If f_{ST} admits continuous second-order partial derivatives, expressions for the bias and the variance of the ideal, infeasible estimator $\hat{f}_{ST}^*$ (2.4), as well as its asymptotic normality, are well known [51], Chapter 4. Proposition 3.1 below ascertains that using the pseudo-observations (2.5)

instead of genuine ones does not affect those properties. Note that (3.2) can be written

$$\hat{f}_{ST}(s,t) = \frac{1}{h^2} \int\int_{\mathbb{R}^2} \phi\left(\frac{s - \Phi^{-1}(u)}{h}\right) \phi\left(\frac{t - \Phi^{-1}(v)}{h}\right) d\hat{C}_n(u,v), \qquad (3.3)$$

where $\hat{C}_n$ is the empirical copula

$$\hat{C}_n(u,v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i \leq u, \hat{V}_i \leq v\}}. \qquad (3.4)$$

Hence, the behavior of $\hat{f}_{ST}(s,t)$ will mostly be driven by the properties of $\hat{C}_n$ on $\mathcal{I}$. Assume the following.

Assumption 3.1. *The sample $\{(X_i, Y_i)\}_{i=1}^{n}$ is an i.i.d. sample from the joint distribution F_{XY}, a bivariate distribution with continuous marginals F_X and F_Y.*

Assumption 3.2. *The copula C of F_{XY} is such that $(\partial C/\partial u)(u,v)$ and $(\partial^2 C/\partial u^2)(u,v)$ exist and are continuous on $\{(u,v) : u \in (0,1), v \in [0,1]\}$, and $(\partial C/\partial v)(u,v)$ and $(\partial^2 C/\partial v^2)(u,v)$ exist and are continuous on $\{(u,v) : u \in [0,1], v \in (0,1)\}$. In addition, there are constants K_1 and K_2 such that*

$$\begin{cases} \left|\dfrac{\partial^2 C}{\partial u^2}(u,v)\right| \leq \dfrac{K_1}{u(1-u)}, & \text{for } (u,v) \in (0,1) \times [0,1]; \\[3mm] \left|\dfrac{\partial^2 C}{\partial v^2}(u,v)\right| \leq \dfrac{K_2}{v(1-v)}, & \text{for } (u,v) \in [0,1] \times (0,1). \end{cases}$$

Assumption 3.3. *The density c of C exists, is positive and admits continuous second-order partial derivatives on the interior of the unit square $\mathcal{I}$. In addition, there is a constant K_{00} such that*

$$c(u,v) \leq K_{00} \min\left(\frac{1}{u(1-u)}, \frac{1}{v(1-v)}\right) \qquad \forall (u,v) \in (0,1)^2. \qquad (3.5)$$

Assumption 3.1 guarantees the existence and the uniqueness of the copula C of F_{XY}. Assumptions 3.2–3.3 mostly reduce to Conditions 2.1 and 4.1 in [47], who claims that they are not restrictive. In particular, they were designed for relaxing to a large extent some of the assumptions previously commonplace in the copula literature, but which were violated by most of the usual copula models used in practice. Specifically, Segers [47] shows that they hold for many copula families, such as Gaussian and extreme-value copulas (subject to some conditions on their Pickands dependence function), and Omelka, Gijbels and Veraverbeke [43] explicitly show that they are satisfied for some common Archimedean copulas such as Clayton and Gumbel, and for Student copulas as well. Compared to [47], Assumption 3.3 only requires further the existence and continuity of second-order partial derivatives of c, which is natural in kernel estimation. It is worth noting that c is allowed to grow unboundedly in some of the corners of $\mathcal{I}$, provided (3.5) remains valid.

Proposition 3.1. *Assume that $\mathbf{K}(z_1, z_2) = \phi(z_1)\phi(z_2)$ and $\mathbf{H}_{ST} = h^2\mathbf{I}$ with $h \sim n^{-a}$ for some $a \in [1/6, 1/4)$. Under Assumptions 3.1–3.3, the estimator (3.2) at any $(s, t) \in \mathbb{R}^2$ is such that*

$$\sqrt{nh^2}\big(\hat{f}_{ST}(s, t) - f_{ST}(s, t) - h^2 b_{ST}(s, t)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_{ST}^2(s, t)\big), \tag{3.6}$$

where $b_{ST}(s, t) = \frac{1}{2}(\frac{\partial^2 f_{ST}}{\partial s^2}(s, t) + \frac{\partial^2 f_{ST}}{\partial t^2}(s, t))$ and $\sigma_{ST}^2(s, t) = \frac{f_{ST}(s,t)}{4\pi}$.

An important observation is that this result holds true for $h \sim n^{-a}$, $a \in [1/6, 1/4)$, which includes the optimal bandwidth order $h \sim n^{-1/6}$ for bivariate density estimation. Now, as recalled in Section 1, resorting to pseudo-observations is known to usually affect the properties of most estimators in copula modeling. In particular, an overriding result in the field is the weak convergence of the empirical copula process

$$\begin{aligned}
\mathbb{C}_n(u, v) &\doteq \sqrt{n}\big(\hat{C}_n(u, v) - C(u, v)\big) \rightsquigarrow \mathbb{G}_C(u, v) \\
&\doteq \mathbb{B}_C(u, v) - \frac{\partial C}{\partial u}(u, v)\mathbb{B}_C(u, 1) - \frac{\partial C}{\partial v}(u, v)\mathbb{B}_C(1, v),
\end{aligned} \tag{3.7}$$

where $\mathbb{B}_C(u, v)$ is the tight centered Gaussian process whose covariance function is $\mathbb{E}(\mathbb{B}_C(u, v)\mathbb{B}_C(u', v')) = C(u \wedge u', v \wedge v') - C(u, v)C(u', v')$ [19,47]. In fact, $\mathbb{B}_C(u, v)$ would be the limiting process if the margins were known, i.e. if true U_i's and V_i's were used in (3.4). The extra two terms in the right-hand side of (3.7) are, therefore, often interpreted as 'the price to pay' for using pseudo-observations – although this effect may be advantageous [27]. Yet, the proof of Proposition 3.1 reveals that the effect of those two terms asymptotically vanishes within (3.3). As a result, the rate of convergence, as well as the asymptotic bias and variance, are the same as those of the ideal estimator $\hat{f}_{ST}^*$ using true (U_i, V_i)'s. This is because a kernel density estimator converges slower than an empirical distribution function. Resorting to pseudo-observations may disturb the $\sqrt{n}$-convergence of the latter, but it goes unnoticed compared to the nonparametric convergence rate $O((nh^2)^{-1/2})$ of the former.

Now, differentiating (2.1) yields

$$\begin{cases}
\dfrac{\partial f_{ST}}{\partial s}(s, t) = \dfrac{\partial c}{\partial u}\big(\Phi(s), \Phi(t)\big)\phi^2(s)\phi(t) - sc\big(\Phi(s), \Phi(t)\big)\phi(s)\phi(t), \\[2mm]
\dfrac{\partial^2 f_{ST}}{\partial s^2}(s, t) = \dfrac{\partial^2 c}{\partial u^2}\big(\Phi(s), \Phi(t)\big)\phi^3(s)\phi(t) - 3s\dfrac{\partial c}{\partial u}\big(\Phi(s), \Phi(t)\big)\phi^2(s)\phi(t) \\[2mm]
\qquad\qquad\quad + \big(s^2 - 1\big)c\big(\Phi(s), \Phi(t)\big)\phi(s)\phi(t)
\end{cases} \tag{3.8}$$

(and similar for $\frac{\partial f_{ST}}{\partial t}$, $\frac{\partial^2 f_{ST}}{\partial t^2}$ and $\frac{\partial^2 f_{ST}}{\partial s\, \partial t}$). Hence, combining (2.1), (2.3), (3.6) and (3.8), one can state:

Theorem 3.1. *Under the assumptions of Proposition 3.1, the 'naive' probit transformation kernel copula density estimator (3.1) at any $(u, v) \in (0, 1)^2$ is such that*

$$\sqrt{nh^2}\big(\hat{c}^{(\tau)}(u, v) - c(u, v) - h^2 b(u, v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma^2(u, v)\big),$$

where

$$b(u, v) = \frac{1}{2}\left\{ \frac{\partial^2 c}{\partial u^2}(u, v)\phi^2(\Phi^{-1}(u)) + \frac{\partial^2 c}{\partial v^2}(u, v)\phi^2(\Phi^{-1}(v)) \right.$$

$$- 3\left(\frac{\partial c}{\partial u}(u, v)\Phi^{-1}(u)\phi(\Phi^{-1}(u)) + \frac{\partial c}{\partial v}(u, v)\Phi^{-1}(v)\phi(\Phi^{-1}(v)) \right) \quad (3.9)$$

$$\left. + c(u, v)\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2 \right) \right\}$$

and $\sigma^2(u, v) = \frac{c(u,v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$.

When (u, v) approaches one of the boundaries, both the (asymptotic) bias and variance of the estimator tend to grow unboundedly. Indeed, $\sigma^2(u, v) \propto c(u, v)/(\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v)))$ and $b(u, v)$ includes the term $c(u, v)(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2)$, and the functions $\Phi^{-1}(\cdot)$ and $1/\phi(\Phi^{-1}(\cdot))$ are unbounded. Thus, along the boundaries, $\hat{c}^{(\tau)}$ will work properly only over areas, if any, where c approaches 0 very smoothly. Otherwise, $\hat{c}^{(\tau)}$ will typically show a very erratic behavior (large variance) and will be prone to exploding (large positive bias), especially in the corners. Figure 1 illustrates these problems, from a typical sample of size $n = 1000$ drawn from the Gaussian copula with correlation $\rho = 0.3$ (left panel). The corresponding naive probit-transformation kernel estimator is shown in the middle panel. An unconstrained matrix $\mathbf{H}_{ST}$ was used in (2.6)–(2.7) and chosen by the multivariate Normal Reference rule [7]. Here, this is optimal: C being a Gaussian copula, f_{ST} is a bivariate normal density. Over the middle of $\mathcal{I}$, the estimator works decently, but towards the boundaries the estimate shows coarse folds and, indeed, hypertrophies the peaks at $(0, 0)$ an $(1, 1)$. Clearly, this estimator is not acceptable as-is. It is, therefore, not surprising that it has been reported not to perform well, see, for example, Bouezmarni, El Ghouch and Taamouti's [4] simulations.
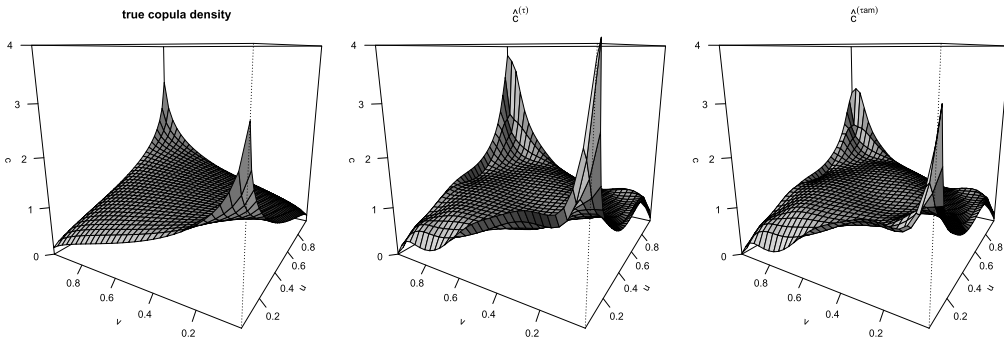

Figure 1. True Gaussian copula density with $\rho = 0.3$ (left), its naive probit-transformation kernel estimator from a typical random sample of size $n = 1000$ (middle) and its amended naive probit-transformation kernel estimator from the same sample (right). The (unconstrained) bandwidth matrix $\mathbf{H}_{ST}$ was chosen by the Normal Reference rule in the (S, T)-domain.

The third, unbounded term in (3.9) can, however, be easily adjusted for. Instead of (2.3), take

$$
\hat{c}^{(\tau\,\mathrm{am})}(u, v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}
$$
$$
\times \frac{1}{1 + (h^2/2)(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2)}.
\tag{3.10}
$$

For this 'amended' version of $\hat{c}^{(\tau)}$, one can see that the asymptotic bias becomes proportional to

$$
b^{(\mathrm{am})}(u, v) = \frac{1}{2}\left\{ \frac{\partial^2 c}{\partial u^2}(u, v)\phi^2(\Phi^{-1}(u)) + \frac{\partial^2 c}{\partial v^2}(u, v)\phi^2(\Phi^{-1}(v)) \right.
$$
$$
\left. - 3\left(\frac{\partial c}{\partial u}(u, v)\Phi^{-1}(u)\phi(\Phi^{-1}(u)) + \frac{\partial c}{\partial v}(u, v)\Phi^{-1}(v)\phi(\Phi^{-1}(v)) \right) \right\}.
$$

In fact, the deterministic, multiplicative amendment in (3.10) exactly makes it up for the third term in (3.9) in the asymptotic development, given that $(1 + h^2)^{-1} = 1 - h^2 + o(h^2)$ as $h \to 0$. The improvement is illustrated in Figure 1 (right panel), where $\hat{c}^{(\tau\,\mathrm{am})}$ for the same data set as in the middle panel is shown. The peaks at $(0, 0)$ and $(1, 1)$ are now roughly of the right height. The wiggly appearance of the estimate along boundaries mostly remains, though, as the (asymptotic) variance is not affected by the amendment. On a side note, the amendment implies that the estimator $\hat{c}^{(\tau\,\mathrm{am})}(u, v)$ does not integrate to 1 over $\mathcal{I}$ any more, which calls for a renormalization such as $\hat{c}^{(\tau\,\mathrm{am})}(u, v) \leftarrow \hat{c}^{(\tau\,\mathrm{am})}(u, v)/\iint_{\mathcal{I}} \hat{c}^{(\tau\,\mathrm{am})}(u, v)\,du\,dv$. This is, however, frequent in other nonparametric density estimation procedures, and is not really a problem.

3.2. Improved probit-transformation kernel copula density estimators

Now the properties of the 'improved' probit-transformation kernel copula density estimators are derived. Again, the results are stated for **K** a product of two univariate Gaussian kernels and $\mathbf{H}_{ST} = h^2\mathbf{I}$, for some $h > 0$, in (2.11). The first version estimates the joint density f_{ST} by the local log-linear estimator $\tilde{f}_{ST}^{(1)}$. Consider first the 'ideal' version $\tilde{f}_{ST}^{*(1)}$ of this estimator, using the true sample $\{(S_i, T_i)\}_{i-1}^n$. From [39], one gets, for all $(s, t) \in \mathbb{R}^2$ at which $f_{ST}(s, t)$ is positive and admits continuous second-order partial derivatives,

$$
\sqrt{nh^2}\big(\tilde{f}_{ST}^{*(1)}(s, t) - f_{ST}(s, t) - h^2 b_{ST}^{(1)}(s, t)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_{ST}^{(1)\,2}(s, t)\big),
\tag{3.11}
$$

where

$$
b_{ST}^{(1)}(s, t) = \frac{1}{2}\left\{ \left(\frac{\partial^2 f_{ST}}{\partial s^2} + \frac{\partial^2 f_{ST}}{\partial t^2} \right)(s, t) - \frac{1}{f_{ST}(s, t)}\left(\left\{ \frac{\partial f_{ST}}{\partial s} \right\}^2 + \left\{ \frac{\partial f_{ST}}{\partial t} \right\}^2 \right)(s, t) \right\}
$$

and $\sigma_{ST}^{(1)\,2}(s,t) = \frac{f_{ST}(s,t)}{4\pi}$. If $f_{ST}(s,t) = 0$ at some (s,t), the singularity of the log-density cannot be accurately approximated by (2.8), but this is ruled out here by Assumption 3.3 which requires c to be positive all over the interior of the unit square. By (2.1), this implies that f_{ST} is positive over $\mathbb{R}^2$.

Define the 'ideal' local log-linear probit-transformation kernel copula density estimator $\tilde{c}^{*(\tau,1)}(u,v) = \tilde{f}_{ST}^{*(1)}(\Phi^{-1}(u), \Phi^{-1}(v))/(\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v)))$. Using (2.1), (2.3) and (3.8) in (3.11), one obtains

$$\sqrt{nh^2}\big(\tilde{c}^{*(\tau,1)}(u,v) - c(u,v) - h^2 b^{(1)}(u,v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma^{(1)\,2}(u,v)\big), \qquad (3.12)$$

where

$$b^{(1)}(u,v) = \frac{1}{2}\bigg\{ \frac{\partial^2 c}{\partial u^2}(u,v)\phi^2\big(\Phi^{-1}(u)\big) + \frac{\partial^2 c}{\partial v^2}(u,v)\phi^2\big(\Phi^{-1}(v)\big)$$

$$- \frac{1}{c(u,v)}\bigg(\Big\{\frac{\partial c}{\partial u}(u,v)\Big\}^2 \phi^2\big(\Phi^{-1}(u)\big) + \Big\{\frac{\partial c}{\partial v}(u,v)\Big\}^2 \phi^2\big(\Phi^{-1}(v)\big)\bigg) \qquad (3.13)$$

$$- \bigg(\frac{\partial c}{\partial u}(u,v)\Phi^{-1}(u)\phi\big(\Phi^{-1}(u)\big) + \frac{\partial c}{\partial v}(u,v)\Phi^{-1}(v)\phi\big(\Phi^{-1}(v)\big)\bigg) - 2c(u,v)\bigg\}$$

and $\sigma^{(1)\,2}(u,v) = \frac{c(u,v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$.

The next result ascertains that, like for the 'naive' estimator, the asymptotic properties of $\tilde{c}^{(\tau,1)}$ are not affected by using pseudo-observations, and are consequently identical to those of the ideal version $\tilde{c}^{*(\tau,1)}$.

Theorem 3.2. *Under the assumptions of Proposition* 3.1, *the 'improved' local log-linear probit-transformation kernel copula density estimator* $\tilde{c}^{(\tau,1)}$ *at any* $(u,v) \in (0,1)^2$ *is such that*

$$\sqrt{nh^2}\big(\tilde{c}^{(\tau,1)}(u,v) - c(u,v) - h^2 b^{(1)}(u,v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma^{(1)\,2}(u,v)\big),$$

where $b^{(1)}(u,v)$ *and* $\sigma^{(1)\,2}(u,v)$ *are given above.*

The variance is the same as that of the 'naive' estimator, but the bias is significantly different. It is now automatically free from any unbounded terms. In fact, Hjort and Jones [31] showed (their expression (7.3)) that the local log-linear estimator $\tilde{f}_{ST}^{(1)}$ and the standard kernel estimator $\hat{f}_{ST}$ satisfy, for all (s,t),

$$\tilde{f}_{ST}^{(1)}(s,t) = \hat{f}_{ST}(s,t)\exp\bigg\{-\frac{1}{2}h^2\bigg[\bigg(\frac{\partial \hat{f}_{ST}(s,t)/\partial s}{\hat{f}_{ST}(s,t)}\bigg)^2 + \bigg(\frac{\partial \hat{f}_{ST}(s,t)/\partial t}{\hat{f}_{ST}(s,t)}\bigg)^2\bigg]\bigg\}. \qquad (3.14)$$

This shows that $\tilde{f}_{ST}^{(1)}$ improves on $\hat{f}_{ST}(s,t)$ by adjusting for the local slopes. From (2.3), (2.11) and an analogue of (3.8) for hat versions, one obtains a similar result for the copula density

estimators:

$$\tilde{c}^{(\tau,1)}(u,v) = \hat{c}(u,v)$$

$$\times \exp\left\{-\frac{1}{2}h^2\left[\left(\frac{\partial\hat{c}(u,v)/\partial u}{\hat{c}(u,v)}\right)^2\phi^2\big(\Phi^{-1}(u)\big) + \left(\frac{\partial\hat{c}(u,v)/\partial v}{\hat{c}(u,v)}\right)^2\phi^2\big(\Phi^{-1}(v)\big)\right.\right.$$

$$\left.-2\left\{\left(\frac{\partial\hat{c}(u,v)/\partial u}{\hat{c}(u,v)}\right)\Phi^{-1}(u)\phi\big(\Phi^{-1}(u)\big) + \left(\frac{\partial\hat{c}(u,v)/\partial v}{\hat{c}(u,v)}\right)\Phi^{-1}(v)\phi\big(\Phi^{-1}(v)\big)\right\}\right.$$

$$\left.+\left\{\Phi^{-1}(u)\right\}^2 + \left\{\Phi^{-1}(v)\right\}^2\right]\bigg\}.$$

So, not only the local log-linear estimator $\tilde{c}^{(\tau,1)}$ adjusts for the slopes of c like in (3.14), it actively acts on the boundary behavior as well. Given that $\phi^2(\Phi^{-1}(\cdot))$ and $\Phi^{-1}(\cdot)\phi(\Phi^{-1}(\cdot))$ tend to 0 towards 0 and 1, the first four terms in the bracket will have little influence towards the boundaries (provided c does not tend to 0 too sharply there). On the other hand, $\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2$ tends to $+\infty$ very fast along boundaries (and all the more in the corners), hence $\hat{c}(u,v)$ is multiplied by something quickly tending to 0 there and this prevents it from exploding. This is similar to what the amendment in (3.10) attempted, but is now automatic. Figure 2 (middle panel) shows the estimate $\tilde{c}^{(\tau,1)}$ for the data set used in Figure 1. It used the cross-validation criterion discussed in Section 4 to select the matrix $\mathbf{H}_{ST}$ in (2.11).

The second improved probit-transformation estimator is obtained when taking $p = 2$ in (2.11). Again, consider first the 'ideal' estimator $\tilde{f}_{ST}^{*(2)}$, computed on the true sample $\{(S_i, T_i)\}_{i=1}^n$. Locally fitting a polynomial of a higher degree is known to reduce the asymptotic bias of the estimator, here from order $O(h^2)$ to order $O(h^4)$ [31,39], sufficient smoothness of f_{ST} permitting. Specifically, if f_{ST} admits continuous fourth-order partial derivatives and is positive at (s,t),
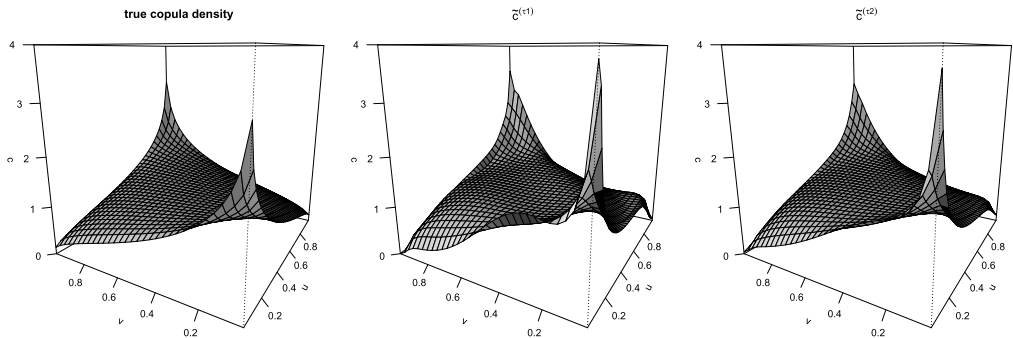

Figure 2. True Gaussian copula density with $\rho = 0.3$ (left), its local log-linear (middle) and log-quadratic (right) improved probit-transformation kernel estimators from the same sample ($n = 1000$) as in Figure 1. Both estimates use an unconstrained bandwidth matrix $\mathbf{H}_{ST}$ chosen by cross-validation in the (S, T)-domain, see Section 4.

then

$$\sqrt{nh^2}\big(\tilde{f}_{ST}^{*(2)}(s,t) - f_{ST}(s,t) - h^4 b_{ST}^{(2)}(s,t)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_{ST}^{(2)\,2}(s,t)\big), \qquad (3.15)$$

where $\sigma_{ST}^{(2)\,2}(s,t) = \frac{5}{2}\frac{f_{ST}(s,t)}{4\pi}$ and

$$b_{ST}^{(2)}(s,t) = -\frac{1}{8} f_{ST}(s,t)\left\{\left(\frac{\partial^4 g}{\partial s^4} + \frac{\partial^4 g}{\partial t^4}\right) + 4\left(\frac{\partial^3 g}{\partial s^3}\frac{\partial g}{\partial s} + \frac{\partial^3 g}{\partial t^3}\frac{\partial g}{\partial t} + \frac{\partial^3 g}{\partial s^2 \partial t}\frac{\partial g}{\partial t} + \frac{\partial^3 g}{\partial s \partial t^2}\frac{\partial g}{\partial s}\right)\right.$$

$$\left. + 2\frac{\partial^4 g}{\partial s^2 \partial t^2}\right\}(s,t),$$

with $g(s,t) = \log f_{ST}(s,t)$. Starting from $g(s,t) = \log c(\Phi(s), \Phi(t)) + \log\phi(s) + \log\phi(t)$, tedious algebraic differentiation provides all partial derivatives of g up to order four in terms of c and its partial derivatives up to order four. Naturally, c will be assumed to admit continuous fourth-order partial derivatives.

Assumption 3.4. *The copula density $c(u,v) = (\partial^2 C/\partial u\,\partial v)(u,v)$ admits continuous fourth-order partial derivatives on the interior of the unit square $\mathcal{I}$.*

As previously, it readily follows from (3.15) that

$$\sqrt{nh^2}\big(\tilde{c}^{*(\tau,2)}(u,v) - c(u,v) - h^4 b^{(2)}(u,v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma^{(2)\,2}(u,v)\big),$$

where $\sigma^{(2)\,2}(u,v) = \frac{5}{2}\frac{c(u,v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$ and $b^{(2)}(u,v)$ is an expression of the same type as (3.13), but involving the partial derivatives of c up to the fourth order. Again, resorting to pseudo-observations will not affect these properties, under a condition on the bandwidth slightly stronger than previously. However, given that the bias order is reduced to $O(h^4)$, the optimal bandwidth order is now seen to be $h \sim n^{-1/10}$, so that the bandwidth requirement does still include that optimal order.

Theorem 3.3. *Under the assumptions of Proposition 3.1 and Assumption 3.4, if $h \sim n^{-a}$ with $a \in [1/10, 1/6]$ as $n \to \infty$, the 'improved' local log-quadratic probit-transformation kernel copula density estimator $\tilde{c}^{(\tau,2)}$ at any $(u,v) \in (0,1)^2$ is such that*

$$\sqrt{nh^2}\big(\tilde{c}^{(\tau,2)}(u,v) - c(u,v) - h^4 b^{(2)}(u,v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma^{(2)\,2}(u,v)\big),$$

where $b^{(2)}(u,v)$ and $\sigma^{(2)\,2}(u,v)$ are described above.

The asymptotic variance is the same as that for $\hat{c}^{(\tau)}(u,v)$ and $\tilde{c}^{(\tau,1)}(u,v)$, except that it is inflated by a factor $5/2$. This inflation factor is a well-known feature when locally fitting a higher-degree polynomial [17], Section 3.3.1. The expression of $b^{(2)}(u,v)$ is not given explicitly here, as it is made up of several dozens of terms. All of them, though, are proportional

to $\{\Phi^{-1}(u)\}^{\alpha}\{\Phi^{-1}(v)\}^{\beta}\phi^{\gamma}(\Phi^{-1}(u))\phi^{\delta}(\Phi^{-1}(v))$, for some non-negative integers $\alpha, \beta, \gamma, \delta$. As these functions all tend to 0 as $u, v \to 0/1$, $b^{(2)}(u, v)$ may actually tend to 0 towards the boundaries, and the bias there be of order $o(h^4)$. Again, this will be the case where c does not tend to 0 too sharply when approaching the boundary.

Interestingly, *ad-hoc* techniques for bias reduction from $O(h^2)$ to $O(h^4)$, e.g. higher-order kernels or multiplicative adjustment, have long been an active research topic in kernel estimation [34]. Yet, few of those methods have actually taken hold, owing to interpretability issues (e.g. negative density estimates when using higher-order kernels) or computational burden. Besides, the demonstrated improvement was asymptotic and usually went unnoticed for typical sample sizes. Here, besides fixing boundary issues, combining transformation and local log-quadratic density estimation achieves that bias reduction with no real extra computational complications. These improvements are visible even in moderately large sample size, as the simulation study in Section 5 will show. A material effect of this is that a larger bandwidth can be used without over-smoothing. This results in smoother estimates, visually more pleasant. This is clear in Figure 2 (right panel), where $\tilde{c}^{(\tau,2)}$ is shown for the same data set as previously. Again, the bandwidth matrix in (2.11) was chosen via the cross-validation method suggested in Section 4.

The above results also justify the choice of the probit transformation in this framework. Basically, the double probit transformation $(S, T) = (\Phi^{-1}(U), \Phi^{-1}(V))$ has for only purpose to send the boundaries of $\mathcal{I}$ away to $\pm\infty$, so that the estimation of f_{ST} is free from boundary issues. Therefore, one could also define $(S, T) = (G_1(U), G_2(V))$ for any two other smooth, monotonic transformations $G_1, G_2 : [0, 1] \to \mathbb{R}$, for example, logit or the quantile function of any $\mathbb{R}$-supported distribution. However, Φ^{-1} transforms the uniformly distributed U and V into normally distributed S and T. These two marginal densities will be estimated with bias of order $o(h^4)$ from the local log-quadratic approximation (2.9), as the 'local parametric model' is right, in the words of Hjort and Jones [31]. Of course, without providing any real guarantee, accurate marginal estimation is an appealing feature that can only be beneficial for estimating the bivariate f_{ST}. In addition, in the particular (but common) case of a Gaussian copula, the probit-transformed density f_{ST} is in fact a bivariate normal density, which can be fully estimated with $o(h^4)$ bias through (2.9). Thus, for Gaussian copula densities (or close thereto), important amount of smoothing may be applied without large bias. It is in consequence of this that the nice, smooth estimate of the Gaussian copula in Figure 2 (right panel) was obtained, see also Figure 3 (right panel) below. Of course, this would not be possible if S and T were not normally distributed, that is, if another transformation than probit was used.

3.3. Improved copula density estimators with *k*-NN bandwidth

Theorems 3.2 and 3.3 reveal that the copula density estimators $\tilde{c}^{(\tau,p)}$, $p = 1, 2$, have advantageous boundary behavior in terms of their bias. However, the fact remains that their variance behaves like

$$\mathrm{Var}\big(\tilde{c}^{(\tau,p)}(u, v)\big) = C_p \frac{c(u, v)}{4\pi nh^2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} + o\big((nh^2)^{-1}\big), \qquad (3.16)$$
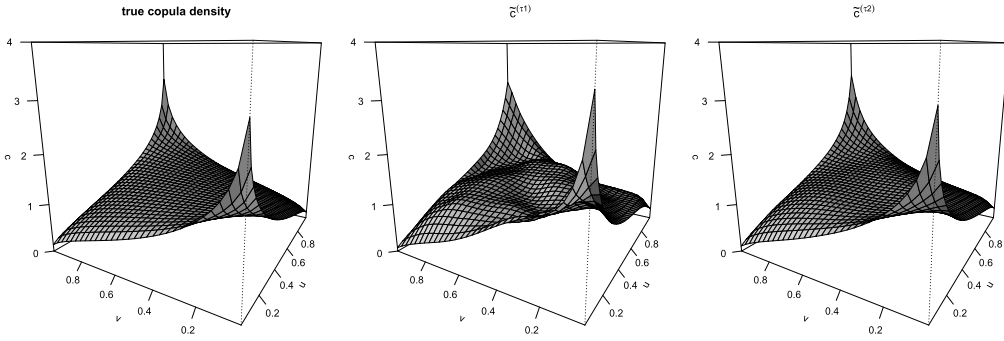
Figure 3. True Gaussian copula density with $\rho = 0.3$ (left), its local log-linear (middle) and log-quadratic (right) improved probit-transformation kernel estimators from the same sample ($n = 1000$) as in Figure 1. Both estimates use an unconstrained bandwidth matrix of type k-NN chosen by cross-validation in the (S, T)-domain, see Section 4.

where $C_1 = 1$ and $C_2 = 5/2$, as $n \to \infty$, growing unboundedly when (u, v) approaches any of the boundaries (except where $c(u, v) \to 0$ very smoothly). Note that this is also the case for other copula density estimators attempting boundary bias correction, see, for instance, [3], Chapter 4 and [32] for similar unbounded boundary variance for the Beta kernel and the Bernstein estimators. In the univariate case, Geenens [22] explained why using a k-Nearest-Neighbor (k-NN) bandwidth in the transformed domain can stabilize the variance of the final estimate towards the boundaries. This idea also appears totally appropriate in the copula setting, which can be understood heuristically as follows.

Again, assume that $\mathbf{H}_{ST}$ in (2.11) is diagonal, but instead of taking $\mathbf{H}_{ST} = h^2 \mathbf{I}$ for some fixed value h, take a local smoothing matrix defined as $\mathcal{H}_{ST}^{(k)}(s, t) = D_k^2(s, t)\mathbf{I}$, where $D_k(s, t)$ is the Euclidean distance between (s, t) and the kth closest observation out of the sample (2.5) in $\mathbb{R}^2$. Now it is k, or equivalently $\alpha = k/n$, that will play the role of the smoothing parameter in lieu of h. If $\mathbf{K}$ had a compact support, α would be the proportion of observations actively entering the estimation of f_{ST} at any (s, t) – this interpretation roughly holds for Gaussian kernels as well. Of course, $D_k(s, t)$ is a random quantity. Following Mack and Rosenblatt [41], one can show that $\mathbb{E}(1/D_k(s, t)) \simeq \frac{\pi f_{ST}(s,t)}{\alpha}$ and, together with $\mathrm{Var}(\tilde{f}_{ST}^{(p)}(s, t)|D_k(s, t)) \simeq C_p \frac{f_{ST}(s,t)}{4\pi n D_k(s,t)}$, that $\mathrm{Var}(\tilde{f}_{ST}^{(p)}(s, t)) \simeq C_p \frac{f_{ST}^2(s,t)}{4n\alpha}$. Now, through (2.3), one directly gets, for all $(u, v) \in (0, 1)^2$,

$$\mathrm{Var}\big(\tilde{c}^{(\tau, p)}(u, v)\big) \simeq C_p \frac{c^2(u, v)}{4n\alpha}.$$

The factor $1/\{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\}$ in (3.16) has been replaced by a factor $c(u, v)$. This is always beneficial to the estimator, given that $c(u, v)\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v)) = f_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))$ goes to 0 as (u, v) approaches a boundary of $\mathcal{I}$, and results in estimates more stable and much smoother in those areas. In fact, given that the (long) tails of $\tilde{f}_{ST}^{(p)}$ in the (S, T)-domain become the (short) boundary regions of $\tilde{c}^{(\tau, p)}$ in $\mathcal{I}$ through the compressing back-transformation

$(u = \Phi(s), v = \Phi(t))$, $\tilde{f}_{ST}^{(p)}$ must have *very smooth* tails in $\mathbb{R}^2$ to produce suitably smooth boundary behavior for $\tilde{c}^{(\tau,p)}$. Local likelihood density estimators using k-NN bandwidth are, indeed, known to produce smoother estimates in the tails than their fixed-bandwidth counterparts, avoiding the occurrence of 'spurious bumps'. Hence, the appropriateness of the method here.

This is illustrated in Figure 3. The two 'improved' probit-transformation kernel copula density estimates (local log-linear, middle; local log-quadratic, right), computed on the same data set as previously, this time use a k-NN-type unconstrained bandwidth matrix $\mathcal{H}_{ST}^{(k)}(s,t)$ (see Section 4 for details). Compared to Figure 2, the estimates are much smoother along the boundaries now, especially $\tilde{c}^{(\tau,2)}$. The value of α selected for the case $p = 1$ was 0.1871, that for the case $p = 2$ was 0.4976: the bias order reduction implied by local log-quadratic modeling allows a larger smoothing parameter to be used. As a result, $\tilde{c}^{(\tau,2)}$ (right panel) has a smooth and visually pleasant appearance, but without oversmoothing. In fact, it is barely distinguishable from the true copula density (left panel). An explanation lies in the last lines of Section 3.2, but the estimator $\tilde{c}^{(\tau,2)}$ with a k-NN-type bandwidth is strikingly good at recovering the shape of any underlying copula density while maintaining a visually pleasant amount of smoothness, see also Section 6.

4. Bandwidth choice

The behavior of kernel estimators crucially depend on their smoothing parameter, whose choice in practice is a very difficult problem especially in more than one dimension. Here an effective way for selecting a suitable bandwidth matrix $\mathbf{H}_{ST}$ in (2.11) is suggested. Importantly, the following methodology does not rely on any copula-specific argument, so can be applied for any bivariate kernel density estimation. Consider

$$\mathbf{H}_{ST} = \begin{pmatrix} h_1^2 & h_{12} \\ h_{12} & h_2^2 \end{pmatrix},$$

an unconstrained bandwidth matrix. The diagonal elements h_1^2 and h_2^2 quantify the amount of smoothing in the directions of the main s- and t-axes, hence their values drive the overall smoothness of the resulting $\tilde{f}_{ST}^{(p)}$ (and eventually that of $\tilde{c}^{(\tau,p)}$). On the other hand, h_{12} sets the direction along which that smoothing mostly applies. For instance, if $\mathbf{K}$ is the bivariate Gaussian kernel, the local weights around $(s,t) \in \mathbb{R}^2$ are set by the elliptical contour lines of the $\mathcal{N}_2((s,t)^t, \mathbf{H}_{ST})$-distribution. If f_{ST} stretches along a particular direction of $\mathbb{R}^2$ (which is the case if c itself does on $\mathcal{I}$), it is greatly beneficial that smoothing be applied in that direction [15], and h_{12} should be selected accordingly. If this is not the case, in particular if S and T are uncorrelated, then h_{12} may be set to 0. This motivates to separate the problem of selecting h_1 and h_2 from that of selecting h_{12}. The idea developed here looks for achieving this, in a way close in spirit to pre-sphering the observations [51], Section 4.6.

Consider the principal components decomposition of the $(n \times 2)$-'data matrix' $\Xi \doteq (\hat{S}_i, \hat{T}_i)_{i=1}^n$. The score of the ith observation on the first and second principal components are given by

$$\hat{Q}_i = W_{11}\hat{S}_i + W_{12}\hat{T}_i, \qquad \hat{R}_i = W_{21}\hat{S}_i + W_{22}\hat{T}_i, \tag{4.1}$$

where $W_1 = (W_{11}, W_{12})^t$ and $W_2 = (W_{21}, W_{22})^t$ are the eigenvectors of $\Xi^T \Xi$. Given that the transformation

$$\begin{pmatrix} Q \\ R \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} S \\ T \end{pmatrix} \doteq \mathbf{W} \begin{pmatrix} S \\ T \end{pmatrix} \tag{4.2}$$

is only a linear reparametrization of $\mathbb{R}^2$, an estimate of f_{ST} can be readily obtained from an estimate of the density of (Q, R), say f_{QR}. In addition, by construction the samples $\{\hat{Q}_i\}$ and $\{\hat{R}_i\}$ are empirically uncorrelated, hence estimating f_{QR} from the sample $\{(\hat{Q}_i, \hat{R}_i)\}_{i=1}^n$ can be based on a diagonal bandwidth matrix $\mathbf{H}_{QR} = \text{diag}(h_Q^2, h_R^2)$ with little side effect. An idea is then to select h_Q and h_R independently via univariate procedures. Denote $\tilde{f}_Q^{(p)}$ and $\tilde{f}_R^{(p)}$ ($p = 1, 2$), the local log-polynomial estimators for the densities of Q and R, respectively, based on the univariate samples $\{\hat{Q}_i\}$ and $\{\hat{R}_i\}$. Of course, $\tilde{f}_Q^{(p)}$ only requires one bandwidth h_Q and $\tilde{f}_R^{(p)}$ only requires one other h_R. Then, h_Q can be selected via cross-validation [38], Section 5.3.3, as

$$h_Q = \arg\min_{h>0} \left\{ \int_{-\infty}^{\infty} \{\tilde{f}_Q^{(p)}(q)\}^2 \, dq - \frac{2}{n} \sum_{i=1}^n \tilde{f}_{Q(-i)}^{(p)}(\hat{Q}_i) \right\}, \tag{4.3}$$

where $\tilde{f}_{Q(-i)}^{(p)}$ is the 'leave-one-out' version of $\tilde{f}_Q^{(p)}$ computed on all the observations but $\hat{Q}_i$. Then h_R can be found similarly, and h_Q and h_R can be plugged into $\mathbf{H}_{QR}$ for proceeding to bivariate estimation. However, optimal bandwidths in one dimension are usually smaller than those for bivariate density estimation of f_{QR}. For the case $p = 1$, the asymptotic optimal bandwidth order is $n^{-1/5}$ for univariate density estimation and $n^{-1/6}$ in two dimensions. For the case $p = 2$, the asymptotic optimal bandwidth order is $n^{-1/9}$ for univariate density estimation and $n^{-1/10}$ in two dimensions. Hence, a fair choice of bandwidth matrix for estimating f_{QR} is $\mathbf{H}_{QR} = K_n^{(p)} \text{diag}(h_Q^2, h_R^2)$, with h_Q and h_R the two bandwidths found above by (univariate) cross-validation, and $K_n^{(1)} = n^{1/15}$ in the local log-linear case and $K_n^{(2)} = n^{1/45}$ in the local log-quadratic case. The estimate of f_{ST} can finally be obtained by linear back-transformation of the estimate of f_{QR}. In fact, due to (4.2), this exactly amounts to directly estimating f_{ST} from $\{(\hat{S}_i, \hat{T}_i)\}$ using the bandwidth matrix $\mathbf{H}_{ST} = K_n^{(p)} \mathbf{W}^{-1} \text{diag}(h_Q^2, h_R^2) \mathbf{W}^{-1}$. There is no need to explicitly estimate f_{QR}.

For a k-NN-type bandwidth matrix, the procedure is very similar. The sample $\{(\hat{Q}_i, \hat{R}_i)\}_{i=1}^n$ is obtained via (4.1). Then, a suitable value of α in the Q-direction is computed as

$$\alpha_Q = \arg\min_{\alpha \in (0,1)} \left\{ \int_{-\infty}^{\infty} \{\tilde{f}_Q^{(p)}(q)\}^2 \, dq - \frac{2}{n} \sum_{i=1}^n \tilde{f}_{Q(-i)}^{(p)}(\hat{Q}_i) \right\}, \tag{4.4}$$

i.e. exactly as in (4.3) but this time $\tilde{f}_Q^{(p)}$ is the estimator based on the k-NN bandwidth α, and so is α_R. Let $\kappa = \alpha_Q/\alpha_R$ and define the squared norm in the (Q, R)-domain as $\|(q, r)\|^2 = q^2 + \kappa^2 r^2$. The factor κ^2 naturally adjusts, through the obtained close-to-optimal values of α_Q and α_R, for the potential discrepancy in local geometry in the q- and r-directions. The bivariate estimation of f_{QR} at any $(q, r) \in \mathbb{R}^2$ is carried out using the $k = K_n^{(p)} \times \alpha_Q \times n$ nearest neighbors of (q, r),

these being determined by the distance implied by the above norm. Here, $K_n^{(1)} = n^{-2/15}$ and $K_n^{(2)} = n^{-4/45}$, again for accounting for the difference in optimal α-orders in one and two dimensions. Finally, the estimate of f_{ST} is obtained by inverse linear transformation or, as set out in the fixed-bandwidth case, directly from the sample $\{(\hat{S}_i, \hat{T}_i)\}_{i=1}^n$ using an appropriate 'Mahalanobis-like' distance. Here, the employed distance makes use through κ of relevant information in terms of optimal smoothing, not only in terms of the covariance structure of $\{(\hat{S}_i, \hat{T}_i)\}$ like the usual Mahalanobis distance. In this setting, the 'smoothing parameters' vector is, therefore, (α_Q, κ).

Admittedly, this procedure may lack of theoretical support. For instance, it is known that pre-sphering the observations in the process of selecting the bandwidth matrix is justified only if (S, T) is bivariate normal, that is, in this framework, if c is the Gaussian copula. Likewise, choosing h_Q/α_Q and h_R/α_R independently via univariate procedures is suitable in theory only if Q and R are independent, not only uncorrelated. The correcting factor $K_n^{(p)}$ may also seem like nothing less than a heuristic, *ad-hoc* correction. Having said this, the procedure gives very reliable results, as illustrated in Figures 2 and 3. This will be even more obvious in the simulation study and the real data analysis detailed in the next sections. In addition, being based on a twofold univariate cross-validation optimization problem, the algorithm is more stable numerically than one based on optimizing a full, bivariate cross-validation criterion. This technique seems, therefore, an acceptable choice for selecting the bandwidth matrix in practice.

5. Simulation study

Here, Monte Carlo simulations results are presented to compare the practical behavior of the probit-transformation estimators with that of their main competitors. All computations have been carried out using the R software and its freely available packages. Specifically, 12 estimators were considered:

- the 'mirror reflection' estimator $\hat{c}^{(m)}$ [28]. It remains a common choice for (ostensibly) correcting boundary bias and will, therefore, be taken as benchmark. A first bandwidth matrix was obtained from the 'augmented' data set (made up of $9n$ 'observations' spread over an area 9 times bigger than $\mathcal{I}$) via the Normal reference rule, then the final matrix was obtained by multiplying the former by $(1/9)^{2/3} \simeq 0.23$ for adjusting for the effective sample size and range;
- the 'naive' probit-transformation estimator $\hat{c}^{(\tau)}$ (2.7) and its amended version $\hat{c}^{(\tau\,am)}$, whose idea is exposed in Section 3.1 (for a general, non-diagonal matrix $\mathbf{H}_{ST}$, the amendment takes a slightly more complicated form than (3.10)). The bandwidth matrix $\mathbf{H}_{ST}$ was selected via a direct plug-in method [14] in the transformed domain (S, T);
- the improved probit-transformation estimators $\tilde{c}^{(\tau, 1)}$ and $\tilde{c}^{(\tau, 2)}$, given by (2.11), based on a k-NN-type bandwidth matrix selected via cross-validation as described at the end of Section 4. As already observed in [22] in the univariate case, when based on a fixed-bandwidth matrix these estimators performed a little less well, so the results are not shown here. The optimization problems (2.11) (local log-polynomial estimation of f_{ST}) and (4.4) (k-NN bandwidth selection) were solved using the relevant functions of the R package locfit. An R package directly implementing these improved probit-transformation estimators is in preparation;

- the Beta kernel estimator $\hat{c}^{(\beta)}$ [8], with Chen's [9] further bias correction. Two smoothing parameters were considered: $h = 0.02$ ($\hat{c}_1^{(\beta)}$) and $h = 0.05$ ($\hat{c}_2^{(\beta)}$);
- the Bernstein copula density estimator $\hat{c}^{(B)}$ [4,5]. Two smoothing parameters were considered: $k = 15$ ($\hat{c}_1^{(B)}$) and $k = 30$ ($\hat{c}_2^{(B)}$);
- the penalized hierarchical *B*-splines estimator $\hat{c}^{(p)}$ [35], computed by the function `pen-copula` in the eponymous R package. The vector of penalty coefficients was set to $\lambda = (10, 10)$ ($\hat{c}_1^{(p)}$), $\lambda = (100, 100)$ ($\hat{c}_2^{(p)}$), and $\lambda = (1000, 1000)$ ($\hat{c}_3^{(p)}$). The parameters d and D were set to 4 and 8, following Kauermann, Schellhase and Ruppert's [35] results.

$M = 1000$ independent random samples $\{(U_i, V_i)\}_{i=1}^n$ of sizes $n = 200$, $n = 500$ and $n = 1000$ were generated from each of the following copulas:

- the independence copula (i.e., U_i's and V_i's drawn independently);
- the Gaussian copula, with parameters $\rho = 0.31$, $\rho = 0.59$ and $\rho = 0.81$;
- the Student *t*-copula with 10 degrees of freedom, with parameters $\rho = 0.31$, $\rho = 0.59$ and $\rho = 0.81$;
- the Student *t*-copula with 4 degrees of freedom, with parameters $\rho = 0.31$, $\rho = 0.59$ and $\rho = 0.81$;
- the Frank copula, with parameter $\theta = 1.86$, $\theta = 4.16$ and $\theta = 7.93$;
- the Gumbel copula, with parameter $\theta = 1.25$, $\theta = 1.67$ and $\theta = 2.5$;
- the Clayton copula, with parameter $\theta = 0.5$, $\theta = 1.67$ and $\theta = 2.5$.

For each family of copulas, the considered three values of the parameter roughly correspond to Kendall's τ's equal to 0.2, 0.4 and 0.6. Of course, all the estimations only made use of the pseudo-observations, that is, the normalized ranks of the observations in the initially generated samples $\{U_i\}_{i=1}^n$ and $\{V_i\}_{i=1}^n$.

The fit of an estimator $\hat{c}$ to a given copula density c is assessed by the Mean Integrated L^2-Error $\mathbb{E}(\int\int_{\mathcal{I}}(\hat{c}(u, v) - c(u, v))^2 \, du \, dv)$, estimated by the average over the 1000 Monte Carlo replications of

$$ISE(\hat{c}) \simeq \frac{1}{(N+1)^2} \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \left\{ \hat{c}\left(\frac{k_1}{N+1}, \frac{k_2}{N+1}\right) - c\left(\frac{k_1}{N+1}, \frac{k_2}{N+1}\right) \right\}^2$$

with $N = 64$. The approximated MISE can be found in Table 1 for $n = 1000$. The results for smaller sample sizes show exactly the same pattern, and are available upon request. For ease of reading and interpretation, all the values are relative to the (approximated) MISE of the benchmark mirror estimator $\hat{c}^{(m)}$. For reference, the effective MISE of $\hat{c}^{(m)}$ is reported in italics in the last column of the table (which is, therefore, not on the same scale as the other values).

The estimators $\tilde{c}^{(\tau,1)}$ and $\tilde{c}^{(\tau,2)}$ are clearly the best, overall, on this L_2-error criterion. They always dramatically improve on the Beta and Bernstein estimators, and they also do much better than the mirror reflection and the penalized B-splines estimators when the dependence is not close to null. When the dependence is very low, $\hat{c}^{(m)}$ and $\hat{c}^{(p)}$ do better, which can be easily understood. It is known that mirror reflection efficiently deals with boundary effects only when the partial derivatives of c are 0 there ('shoulder'). It is, therefore, particularly appropriate for the independence copula ($c \equiv 1$) and other very flat copula densities such as Gaussian or Frank with

Table 1. (approximated) MISE relative to the MISE of the mirror-reflection estimator (last column), $n = 1000$. Bold values show the minimum MISE for the corresponding copula (non-significantly different values are highlighted as well)

$n = 1000$	$\hat{c}^{(\tau)}$	$\hat{c}^{(\tau \, am)}$	$\tilde{c}^{(\tau,1)}$	$\tilde{c}^{(\tau,2)}$	$\hat{c}_1^{(\beta)}$	$\hat{c}_2^{(\beta)}$	$\hat{c}_1^{(B)}$	$\hat{c}_2^{(B)}$	$\hat{c}_1^{(p)}$	$\hat{c}_2^{(p)}$	$\hat{c}_3^{(p)}$	$\hat{c}^{(m)}$
Indep	3.57	2.80	2.89	1.40	7.96	11.65	1.69	3.43	1.62	0.50	**0.14**	*0.01*
Gauss2	2.03	1.52	1.60	0.76	4.63	6.06	1.10	1.82	0.98	**0.66**	0.89	*0.01*
Gauss4	0.63	0.49	0.44	**0.21**	1.72	1.60	0.75	0.58	0.62	0.99	2.93	*0.05*
Gauss6	0.21	0.20	0.11	**0.05**	0.74	0.33	0.77	0.37	0.72	1.21	2.83	*0.26*
Std(10)2	1.36	1.06	1.04	**0.55**	3.07	3.98	0.96	1.24	0.86	0.87	1.48	*0.02*
Std(10)4	0.41	0.37	0.28	**0.15**	1.22	1.00	0.74	0.46	0.68	1.08	2.51	*0.08*
Std(10)6	0.15	0.18	**0.08**	**0.05**	0.71	0.24	0.79	0.41	0.84	1.21	2.36	*0.39*
Std(4)2	0.61	0.56	0.50	**0.40**	1.57	1.80	0.78	0.67	0.75	1.01	1.88	*0.04*
Std(4)4	0.21	0.27	**0.17**	**0.15**	0.88	0.51	0.75	0.42	0.75	1.12	2.07	*0.16*
Std(4)6	**0.09**	0.17	**0.08**	**0.09**	0.70	0.19	0.82	0.47	0.90	1.17	1.90	*0.67*
Frank2	3.31	2.42	2.57	1.35	7.16	9.63	1.70	2.95	1.31	**0.45**	**0.49**	*0.01*
Frank4	2.35	1.45	1.51	0.99	4.42	4.89	1.49	1.65	**0.60**	0.72	6.14	*0.01*
Frank6	0.96	0.52	**0.45**	**0.44**	1.51	1.19	1.35	0.76	0.65	1.58	7.25	*0.07*
Gumbel2	0.65	0.62	0.56	**0.43**	1.77	1.97	0.82	0.75	0.83	1.03	1.52	*0.04*
Gumbel4	**0.18**	0.28	**0.16**	**0.19**	0.89	0.41	0.78	0.47	0.81	1.10	1.78	*0.21*
Gumbel6	**0.09**	0.21	**0.10**	0.15	0.78	0.29	0.85	0.58	0.94	1.12	1.63	*0.93*
Clayton2	0.63	0.60	0.51	**0.34**	1.78	1.99	0.78	0.70	0.79	1.04	1.79	*0.04*
Clayton4	**0.11**	0.26	**0.10**	**0.15**	0.79	0.27	0.83	0.56	0.90	1.10	1.50	*0.65*
Clayton6	**0.11**	0.28	**0.08**	0.15	0.82	0.35	0.88	0.67	0.96	1.09	1.36	*1.61*

low dependence. The penalized B-splines estimator does even better when using a huge penalty for roughness, for obvious reasons. In all other cases, and particularly when the copula density becomes unbounded in some corners (but not only), $\tilde{c}^{(\tau,1)}$ and $\tilde{c}^{(\tau,2)}$ dramatically outperform their competitors. In fact, mirror reflection and splines are not appropriate methods for estimating unbounded copula densities. By construction, the Beta kernel estimator always tends to be zero along boundaries (see, for instance, Figure 5 below), hence its even worse performance in this framework. The Bernstein estimator does better than $\hat{c}^{(\beta)}$, but cannot really compete with $\tilde{c}^{(\tau,1)}$ and $\tilde{c}^{(\tau,2)}$. Of course, one can argue that the smoothing parameters used for $\hat{c}^{(\beta)}$, $\hat{c}^{(B)}$ and $\hat{c}^{(p)}$ have been selected mostly arbitrarily and are not adequate. This may be true, however, there is no simple, data-driven way of selecting those parameters, hence the choice was made subjectively exactly as a practitioner should have resolved to act. In addition, the above observations support that bad smoothing parameter choice is not the only reason for the poor performance of some of those estimators. Other evidence of that will be given in the next section.

In general, $\tilde{c}^{(\tau,2)}$ is doing better than $\tilde{c}^{(\tau,1)}$, which is expected from the theoretical results. A notable exception, though, is in presence of high tail dependence, i.e. when the copula density tends very quickly to ∞ at one of the corners of $\mathcal{I}$, such as for Clayton and Gumbel copulas with high Kendall's τ. In fact, the extra smoothness guaranteed by local log-quadratic estimation prevents the estimator from growing too quickly in the corners, and this is thus slightly detrimental in those cases. The same comment holds true when comparing the naive estimator
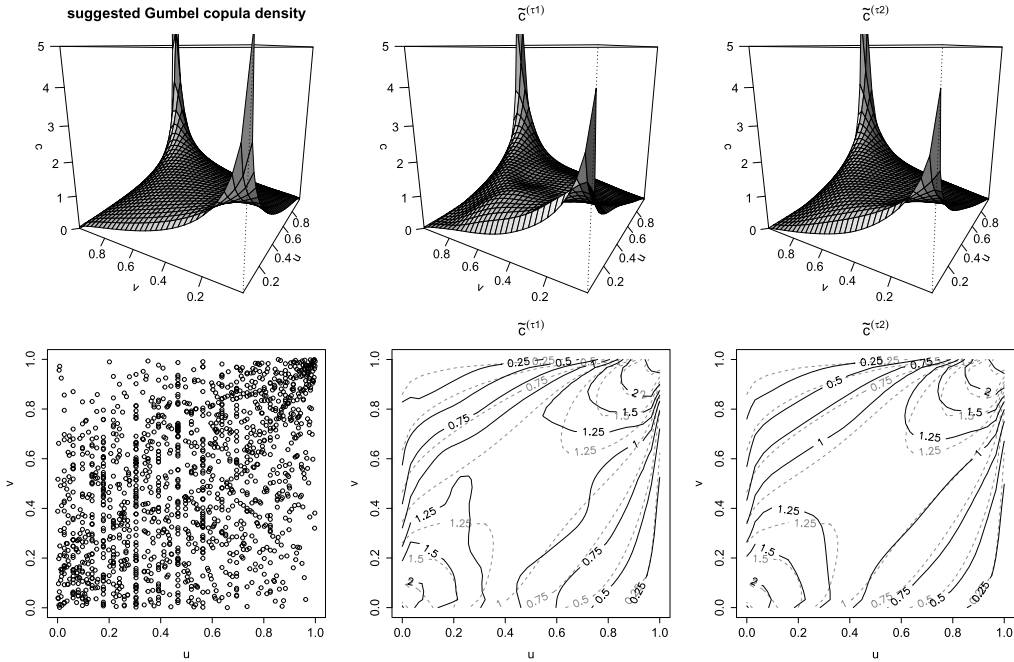
Figure 4. Loss-ALAE dataset: suggested parametric copula density (Gumbel with parameter $\hat{\theta} = 1.453$; upper-left panel) and probit-transformation estimates with $p = 1$ (middle column) and $p = 2$ (right column). The upper line shows 3-d views and the bottom line shows contour lines, superimposed on the Gumbel copula density contour lines. Pseudo-observations are shown in the bottom-left panel.

$\hat{c}^{(\tau)}$ to its amended version $\hat{c}^{(\tau \, \mathrm{am})}$. Generally, $\hat{c}^{(\tau \, \mathrm{am})}$ has lower MISE than $\hat{c}^{(\tau)}$, except in the above-mentioned cases of high tail dependence: the amendment prevents the estimate from exploding, even when the true density does. In any case, these 'naive' versions cannot really match the performance of the 'improved' versions $\tilde{c}^{(\tau,1)}$ and $\tilde{c}^{(\tau,2)}$ on MISE.

6. Real data analysis

In this section, the well-known 'Loss-ALAE' dataset, reporting the indemnity payment (X_i's) and allocated loss adjustment expense (Y_i's) associated to 1500 losses from an insurance company, is considered. Analyzed in [21,36] and [13], this dataset is a classic in the copula literature. In particular, Frees and Valdez [21] mentioned that the Gumbel copula with $\hat{\theta} = 1.453$ provides an excellent fit. The data set initially contains 34 censored observations, that were excluded here as the suggested estimators were not designed to take censorship into account. Using more advanced model selection ideas, Chen *et al.* [11] also found that the Gumbel copula (with the same parameter $\hat{\theta}$) fits the dataset (restricted to its complete cases) the best out of most of the usual parametric copula models. The aim here is to test the probit-transformation estimators $\tilde{c}^{(\tau,p)}$ ($p = 1, 2$) (and their competitors) against that parametric 'gold standard', shown in Figure 4 (up-left).
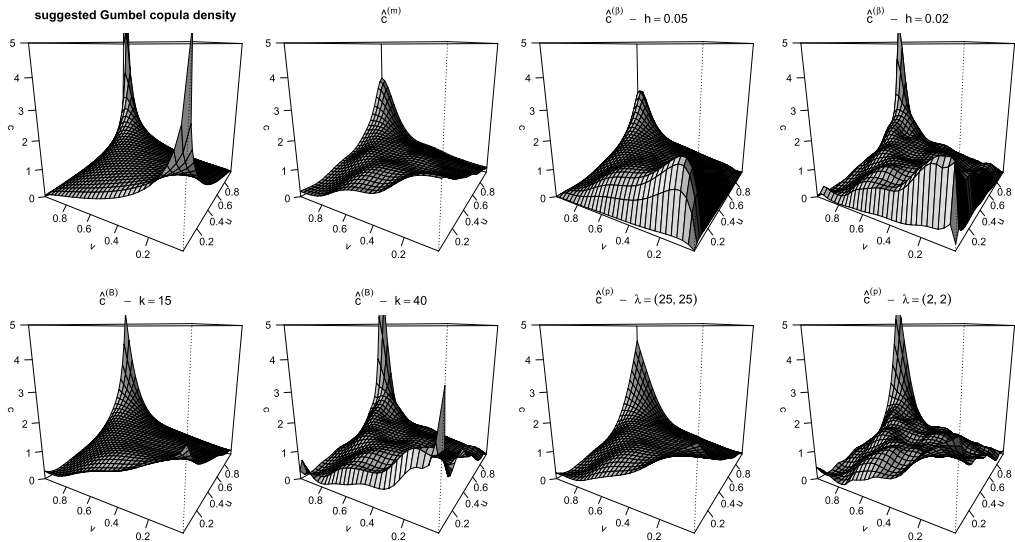
Figure 5. Loss-ALAE dataset: suggested Gumbel copula density ($\hat{\theta} = 1.456$), mirror reflection estimator, Beta kernel estimators with $h = 0.05$ and $h = 0.02$, Bernstein estimators with $k = 15$ and $k = 40$ and penalized B-splines estimators with $d = 4$, $D = 8$ and with $\lambda = (25, 25)$ and $\lambda = (2, 2)$.

The two probit-transformation estimators (local log-linear and local log-quadratic) were first fit to the data set. In both cases, a k-NN bandwidth matrix was used. Using the selection rule prescribed in Section 4, the parameters $(\alpha, \kappa) = (0.24, 1.28)$ for $p = 1$ and $(\alpha, \kappa) = (0.51, 1.01)$ for $p = 2$ were obtained in an automatic manner. The estimator $\tilde{c}^{(\tau,2)}$ is, again, very similar to the parametric fit. In particular, it has that very smooth and pleasant appearance of parametric estimates, while being based on a fully nonparametric procedure. Reproducing 'parametric smoothness' without sacrificing any flexibility is, of course, a substantial achievement. Naturally, $\tilde{c}^{(\tau,1)}$ is less smooth (see discussion in Section 3.2), but is still totally acceptable. Both nonparametric estimates suggest that the true underlying copula density decays towards the $(0, 1)$-corner quicker than what the Gumbel model shows (this is particularly clear from the contour lines). Admittedly, there is no way of knowing what is the truth here. However, $\tilde{c}^{(\tau,1)}$ and $\tilde{c}^{(\tau,2)}$ are based only on the data (it is visually obvious that the upper-left corner of $\mathcal{I}$ is much less endowed in data than the bottom-right corner), and not on any prior assumption. On the contrary, the Gumbel copula density is inherently symmetric in u and v. The peak at $(0, 0)$ also appears less high on the nonparametric estimates than for the Gumbel model.

Figure 5 shows the competitors on the same data set: the mirror reflection estimator, two Beta kernel estimators, two Bernstein estimators and two penalized B-splines estimators. Of course, $\hat{c}^{(m)}$ cannot cope with this unbounded copula density. For the other three methods, producing an estimate reasonably smooth required a value of the smoothing parameter (h for Beta kernel estimators, k for Bernstein estimators and λ for penalized B-splines) preventing correct estimation of the peaks at $(0, 0)$ and $(1, 1)$. To get estimates showing a peak at $(1, 1)$ of roughly the right magnitude, one needed to use smoothing parameters producing unacceptably undersmoothed es-

timation elsewhere on $\mathcal{I}$, yet not even able to properly catch the peak at $(0, 0)$. If the Gumbel copula density is assumed to be close to the truth for this data set, then there is no question that $\tilde{c}^{(\tau,1)}$ and $\tilde{c}^{(\tau,2)}$ are, by far, the best. This, also, illustrates that the results obtained in the simulations are not only due to bad smoothing parameter choices.

7. Concluding remarks

Development of efficient kernel-type methods for nonparametric copula modeling have been delayed owing mainly to the bounded support of copulas, namely the unit square $\mathcal{I}$. It is, indeed, well known that kernel estimators heavily suffer from boundary bias issues, which are not trivial to fix. In this paper, a new kernel-type estimator for the copula density has been proposed. It is based on the probit-transformation idea suggested in [8] and studied in full in the univariate case in [22]. This 'improved probit-transformation estimator' deals with boundary bias in a very natural way. In addition, it has been seen to easily cope with potentially unbounded copula densities, which are the common and interesting cases in copula modeling. An easy-to-implement selection rule for the necessary smoothing parameters has also been proposed. This procedure has been seen to be very stable and to give very good results in practice. In particular, a version of the estimator ($\tilde{c}^{(\tau,2)}$ with k-NN-type bandwidth matrix) is able to reproduce the smooth and pleasant appearance of parametric models, while keeping the flexibility of fully nonparametric estimation procedures. A comprehensive simulation study has emphasized the very good practical performance of that estimator compared to its main competitors. As of now, though, the theoretical properties of the estimator have been derived under the assumption of i.i.d. sampling, making use of the strong approximation for the empirical copula process provided by Proposition 4.2 of [47]. To the best of these authors' knowledge, this result has not been proved in the case of weakly dependent observations. It would be particularly significant to investigate this in a near future, given the predominant place recently found by copula modeling in the setting of time series, notably in finance. Other directions for future research would look for using the new copula density estimator in a variety of problems, for instance copula goodness-of-fit tests [18,46].

Supplementary Material

Supplement to "Probit transformation for nonparametric kernel estimation of the copula density" (DOI: 10.3150/15-BEJ798SUPP; .pdf). An appendix consisting of the proofs of the different theoretical results is available as supplementary material.

References

[1] Autin, F., Le Pennec, E. and Tribouley, K. (2010). Thresholding methods to estimate copula density. *J. Multivariate Anal*. **101** 200–222. MR2557629

[2] Behnen, K., Hušková, M. and Neuhaus, G. (1985). Rank estimators of scores for testing independence. *Statist. Decisions* **3** 239–262. MR0809363

[3] Blumentritt, T. (2011). On copula density estimation and measures of multivariate association. Ph.D. Dissertation, Universität zu Köln.

[4] Bouezmarni, T., El Ghouch, A. and Taamouti, A. (2013). Bernstein estimator for unbounded copula densities. *Stat. Risk Model.* **30** 343–360. MR3143795

[5] Bouezmarni, T., Rombouts, J.V.K. and Taamouti, A. (2010). Asymptotic properties of the Bernstein density copula estimator for α-mixing data. *J. Multivariate Anal.* **101** 1–10. MR2557614

[6] Bücher, A. and Volgushev, S. (2013). Empirical and sequential empirical copula processes under serial dependence. *J. Multivariate Anal.* **119** 61–70. MR3061415

[7] Chacón, J.E., Duong, T. and Wand, M.P. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statist. Sinica* **21** 807–840. MR2829857

[8] Charpentier, A., Fermanian, J.-D. and Scaillet, O. (2007). The estimation of copulas: Theory and practice. In *Copulas: From Theory to Application in Finance* (J. Rank, ed.) 35–60. London: Risk Publications.

[9] Chen, S.X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* **31** 131–145. MR1718494

[10] Chen, S.X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canad. J. Statist.* **35** 265–282. MR2393609

[11] Chen, X., Fan, Y., Pouzo, D. and Ying, Z. (2010). Estimation and model selection of semiparametric multivariate survival functions under general censorship. *J. Econometrics* **157** 129–142. MR2652285

[12] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.* (5) **65** 274–292. MR0573609

[13] Denuit, M., Purcaru, O. and Van Keilegom, I. (2006). Bivariate Archimedean copula modelling for censored data in non-life insurance. *J. Actuar. Pract.* **13** 5–32.

[14] Duong, T. and Hazelton, M.L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr. Stat.* **15** 17–30. MR1958957

[15] Duong, T. and Hazelton, M.L. (2005). Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation. *J. Multivariate Anal.* **93** 417–433. MR2162646

[16] Embrechts, P. (2009). Copulas: A personal view. *J. Risk Insur.* **76** 639–650.

[17] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. London: Chapman & Hall. MR1383587

[18] Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *J. Multivariate Anal.* **95** 119–152. MR2164126

[19] Fermanian, J.-D., Radulović, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860. MR2093613

[20] Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *J. Risk* **5** 25–54.

[21] Frees, E.W. and Valdez, E.A. (1998). Understanding relationships using copulas. *N. Am. Actuar. J.* **2** 1–25. MR1988432

[22] Geenens, G. (2014). Probit transformation for kernel density estimation on the unit interval. *J. Amer. Statist. Assoc.* **109** 346–358. MR3180568

[23] Geenens, G., Charpentier, A. and Paindaveine, D. (2016). Supplement to "Probit transformation for nonparametric kernel estimation of the copula density." DOI:10.3150/15-BEJ798SUPP.

[24] Genest, C., Masiello, E. and Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance Math. Econom.* **44** 170–181. MR2517883

[25] Genest, C. and Rémillard, B. (2004). Tests of independence and randomness based on the empirical copula process. *TEST* **13** 335–370. MR2154005

[26] Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.* **44** 199–213. MR2517885

[27] Genest, C. and Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal*. **101** 1837–1845. MR2651959

[28] Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Comm. Statist. Theory Methods* **19** 445–464. MR1067582

[29] Gudendorf, G. and Segers, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *J. Statist. Plann. Inference* **142** 3073–3085. MR2956794

[30] Härdle, W.K. and Okhrin, O. (2010). De copulis non est disputandum. Copulae: An overview. *AStA Adv. Stat. Anal*. **94** 1–31. MR2609887

[31] Hjort, N.L. and Jones, M.C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist*. **24** 1619–1647. MR1416653

[32] Janssen, P., Swanepoel, J. and Veraverbeke, N. (2014). A note on the asymptotic behavior of the Bernstein estimator of the copula density. *J. Multivariate Anal*. **124** 480–487. MR3147339

[33] Joe, H. (1997). *Multivariate Models and Dependence Concepts. Monographs on Statistics and Applied Probability* **73**. London: Chapman & Hall. MR1462613

[34] Jones, M.C. and Signorini, D.F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Statist. Assoc*. **92** 1063–1073. MR1482137

[35] Kauermann, G., Schellhase, C. and Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical B-splines. *Scand. J. Stat*. **40** 685–705. MR3145112

[36] Klugman, S.A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance Math. Econom*. **24** 139–148. MR1710816

[37] Li, B. and Genton, M.G. (2013). Nonparametric identification of copula structures. *J. Amer. Statist. Assoc*. **108** 666–675. MR3174650

[38] Loader, C. (1999). *Local Regression and Likelihood*. New York: Springer. MR1704236

[39] Loader, C.R. (1996). Local likelihood density estimation. *Ann. Statist*. **24** 1602–1618. MR1416652

[40] Lopez-Paz, D., Hernández-Lobato, J.M. and Schölkopf, B. (2013). Semi-supervised domain adaptation with non-parametric copulas. Unpublished manuscript.

[41] Mack, Y.P. and Rosenblatt, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Multivariate Anal*. **9** 1–15. MR0530638

[42] Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed. New York: Springer. MR2197664

[43] Omelka, M., Gijbels, I. and Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Ann. Statist*. **37** 3023–3058. MR2541454

[44] Qu, L. and Yin, W. (2012). Copula density estimation by total variation penalized likelihood with linear equality constraints. *Comput. Statist. Data Anal*. **56** 384–398. MR2842346

[45] Salmon, F. (2009). Recipe for disaster: The formula that killed wall street. *Wired Magazine*, February 23.

[46] Scaillet, O. (2007). Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters. *J. Multivariate Anal*. **98** 533–543. MR2293013

[47] Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900

[48] Shen, X., Zhu, Y. and Song, L. (2008). Linear B-spline copulas with applications to nonparametric estimation of copulas. *Comput. Statist. Data Anal*. **52** 3806–3819. MR2427382

[49] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229–231. MR0125600

[50] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist*. **33** 357–375. MR2193980

[51] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability* **60**. London: Chapman & Hall. MR1319818