2

The Estimation of Copulas: Theory and Practice

Arthur Charpentier; Jean-David Fermanian; Olivier Scaillet[1]

Ensae-Crest and Katholieke Universiteit Leuven; BNP-Paribas and Crest; HEC Genève and Swiss Finance Institute

INTRODUCTION

Copulas are a way of formalising dependence structures of random vectors. Although they have been known about for a long time (Sklar (1959)), they have been rediscovered relatively recently in applied sciences (biostatistics, reliability, biology, etc). In finance, they have become a standard tool with broad applications: multi-asset pricing (especially complex credit derivatives), credit portfolio modelling, risk management, etc. For example, see Li (1999), Patton (2001) and Longin and Solnik (1995).

Although the concept of copulas is well understood, it is now recognised that their empirical estimation is a harder and trickier task. Many traps and technical difficulties are present, and these are, most of the time, ignored or underestimated by practitioners. The problem is that the estimation of copulas implies usually that every marginal distribution of the underlying random vectors must be evaluated and plugged into an estimated multivariate distribution. Such a procedure produces unexpected and unusual effects with respect to the usual statistical procedures: non-standard limiting behaviours, noisy estimations, etc (eg, see the discussion in Fermanian and Scaillet, 2005).

In this chapter, we focus on the practical issues practitioners are faced with, in particular concerning estimation and visualisation.

In the first section, we give a general setting for the estimation of copulas. Such a framework embraces most of the available techniques. In the second section, we deal with the estimation of the copula density itself, with a particular focus on estimation near the boundaries of the unit square.

A GENERAL APPROACH FOR THE ESTIMATION OF COPULA FUNCTIONS

Copulas involve several underlying functions: the marginal cumulative distribution functions (CDF) and a joint CDF. To estimate copula functions, the first issue consists in specifying how to estimate separately the margins and the joint law. Moreover, some of these functions can be fully known. Depending on the assumptions made, some quantities have to be estimated parametrically, or semi- or even non-parametrically. In the latter case, the practitioner has to choose between the usual methodology of using "empirical counterparts" and invoking smoothing methods well-known in statistics: kernels, wavelets, orthogonal polynomials, nearest neighbours, etc.

Obviously, the estimation precision and the graphical results are functions of all these choices. A true known marginal can greatly improve the results under well-specification, but the reverse is true under misspecification (even under a light one). Without any valuable prior information, non-parametric estimation should be favoured, especially for marginal estimation.

To illustrate this point Figure 2.1 shows the graphical behaviour of the exceeding probability function

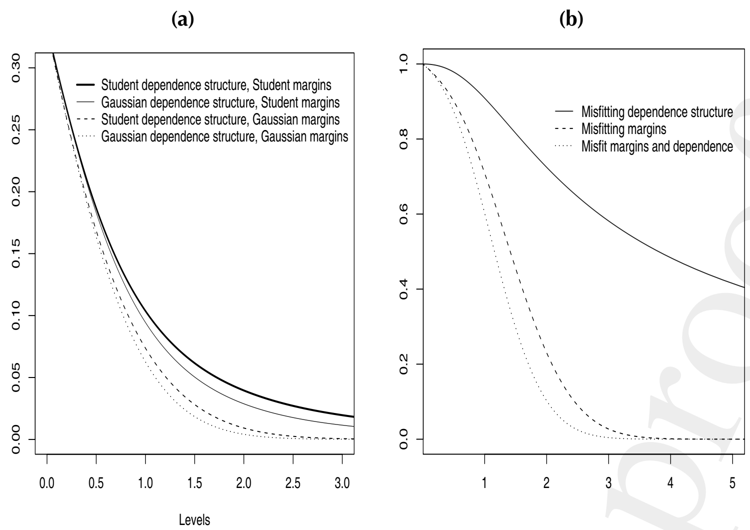
$$\chi : p \mapsto P(X > F_X^{-1}(p), Y > F_Y^{-1}(p))$$

If the true underlying model is a multivariate Student vector (X, Y), the associated probability is the upper line. If either marginal distributions are misspecified (eg, Gaussian marginal distributions), or the dependence structure is misspecified (eg, joint Gaussian distribution), these probabilities are always underestimated, especially in the tails.

Now, let us introduce our framework formally. Consider the estimation of a d-dimensional copula C, that can be written

$$C(\boldsymbol{u}) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))$$

Figure 2.1 (a) The function χ when (X, Y) is a Student random vector, and when either margins or the dependence structure are misspecified. (b) The associated ratios of exceeding probability corresponding to the χ function obtained for the misspecified model versus the true χ (for the true Student model).



Obviously, all the marginal CDFs have been denoted by F_k, $k = 1, \ldots, d$, when the joint CDF is F. Throughout this chapter, the inverse operator $^{-1}$ should be understood to be a generalised inverse; namely that for every function G,

$$G^{-1}(x) = \inf\{y \mid G(y) \geq x\}$$

Assume we have observed a T-sample $(\boldsymbol{X}_i)_{i=1,\ldots,T}$. These are some realisations of the d-random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$. Note that we do not assume that $\boldsymbol{X}_i = (X_{1i}, \ldots, X_{di})$ are mutually independent (at least for the moment).

Every marginal CDF, say the kth, can be estimated empirically by

$$F_k^{(1)}(x) = \frac{1}{T} \sum_{i=1}^{T} \mathbb{1}(X_{ki} \leq x)$$

and $[F_k^{(1)}]^{-1}(u_k)$ is simply the empirical quantile corresponding to $u_k \in [0,1]$. Another means of estimation is to smooth such CDFs, and the simplest way is to invoke the kernel method (eg, see Härdle and Linton (1994) or Pagan and Ullah (1999) for an introduction): consider a univariate kernel function $K : \mathbb{R} \longrightarrow \mathbb{R}$, $\int K = 1$, and a bandwidth sequence h_T (or simply h hereafter), $h_T > 0$ and $h_T \longrightarrow 0$ when $T \rightarrow \infty$. Then, $F_k(x)$ can be estimated by

$$F_k^{(2)}(x) = \frac{1}{T} \sum_{i=1}^{T} \mathbb{K}\left(\frac{x - X_{ki}}{h} \right)$$

for every real number x, by denoting $\mathbb{K}$ the primitive function of K: $\mathbb{K}(x) = \int_{-\infty}^{x} K$.

There exists another common case: assume that an underlying parametric model has been fitted previously for the kth margin. Then, the natural estimator for $F_k(x)$ is some CDF $F_k^{(3)}(x, \hat{\theta}_k)$ that depends on the relevant estimated parameter $\hat{\theta}_k$. When such a model is well-specified, $\hat{\theta}_k$ is tending almost surely to a value θ_k such that $F_k(\cdot) = F_k^{(3)}(\cdot, \theta_k)$. The last limiting case is the knowledge of the true CDF F_k. Formally, we will set $F_k^{(0)} = F_k$.

Similarly, the joint CDF F can be estimated empirically by

$$F^{(1)}(\boldsymbol{x}) = \frac{1}{T} \sum_{i=1}^{T} \mathbb{1}(\boldsymbol{X}_i \leq \boldsymbol{x})$$

or by the kernel method

$$F^{(2)}(\boldsymbol{x}) = \frac{1}{T} \sum_{i=1}^{T} \mathbb{K}\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h} \right)$$

with a d-dimensional kernel K, so that

$$\mathbb{K}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} K$$

for every $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Besides, there may exist an underlying parametric model for $\boldsymbol{X}$: F is assumed to belong to a set of multivariate CDFs indexed by a parameter τ. A consistent estimation $\hat{\tau}$ for the "true" value τ allows setting $F^{(3)}(\cdot) = F(\cdot, \hat{\tau})$. Finally, we can denote $F^{(0)} = F$.

Therefore, generally speaking, a d-dimensional copula C can be estimated by

$$\hat{C}(\boldsymbol{u}) = F^{(j)}\left([F_1^{(j_1)}]^{-1}(u_1), \ldots, [F_d^{(j_d)}]^{-1}(u_d)\right) \tag{2.1}$$

for each of the indexes $j, j_1, j_2, \ldots, j_d$ that belong to $\{0, 1, 2, 3\}$. Thus, it is not so obvious to discriminate between all these competitors, especially without any parametric assumption.

Every estimation method has its own advantages and drawbacks. The full empirical method ($j = j_1 = \cdots = j_d = 1$ with the notations of Equation (2.1)) has been introduced in Deheuvels (1979, 1981a, 1981b) and studied more recently by Fermanian *et al* (2004), in the independent setting, and by Doukhan *et al* (2004) in a dependent framework. It provides a robust and universal way for estimation purposes. Nonetheless, its discontinuous feature induces some difficulties: the graphical representations of the copula can be not very nice from a visual point of view and not intuitive. Moreover, there is no unique choice for building the inverse function of $F_k^{(1)}$. In particular, if $X_{k1} \leq \cdots \leq X_{kT}$ is the ordered sample on the k-axis, the inverse function of $F_k^{(1)}$ at some point i/T may be chosen arbitrarily between X_{ki} and $X_{k(i+1)}$. Finally, since the copula estimator is not differentiable when only one empirical CDF is involved in Equation (2.1), it cannot, for example, be used straightforwardly to derive an estimate of the associated copula density (by differentiation of $\hat{C}(\boldsymbol{u})$ with respect to all its arguments) or for optimisation purposes.

Smooth estimators are better suited to graphical usage, and can provide more easily the intuition to achieve the "true" underlying parametric distribution. However, they depend on an auxiliary smoothing parameter (eg, h in the case of the kernel method), and suffer from the well-known "curse of dimensionality": the higher the dimension (d with our notations), the worse the performance in terms of convergence rates. In other words, as the dimension increases, the complexity of the problem increases exponentially.[2] Such methods can be invoked safely in practice when $d \leq 3$ and for sample sizes larger than, say, two hundred observations (which is usual in finance). The theory of fully smoothed copulas ($j = j_1 = \cdots = j_d = 2$ with the notation of Equation (2.1)) can be found in Fermanian and Scaillet (2003) in a strongly dependent framework.

A more comfortable situation exists when "good" parametric assumptions are put into (2.1) for the marginal CDFs and/or the joint CDF F. The former case is relatively usual because there exist a great many univariate models for financial variables (eg, see Alexander (2002)). Nevertheless, for a lot of dynamic models (eg, stochastic volatility models), their (unconditional) marginal CDFs cannot be written explicitly. Obviously, we are under the threat of a misspecification, which can have disastrous effects (see Fermanian and Scaillet (2005)). Concerning a parametric assumption for F itself, our opinion is balanced. At first glance, we are absolutely free to choose an "interesting" parametric family $\mathcal{F}$ of d-dimensional CDFs that would contain the true law F. But, by setting for every real number x and every $k = 1, \dots, d$

$$F_k^{(3)}(x) = F(+\infty, \dots, +\infty, x, +\infty, \dots \mid \hat{\tau})$$

where x is the kth argument of F, we should have found the "right" marginal distributions too, to be self-coherent. Indeed, the joint law contains the marginal ones. Then, the estimated copula should be

$$\hat{C}(\boldsymbol{u}) = F^{(3)}\left([F_1^{(3)}]^{-1}(u_1), [F_2^{(3)}]^{-1}(u_2), \dots, [F_d^{(3)}]^{-1}(u_d)\right)$$

In reality, the problem is finding a sufficiently rich family $\mathcal{F}$ *ex ante* that might generate all empirical features. What people do is more clever. They choose a parametric family $\mathcal{F}^*$ and other marginal parametric families $\mathcal{F}_k^*$, $k = 1, \dots, d$, and set

$$C(\boldsymbol{u}) = \hat{F}^*\left([\hat{F}_1^*]^{-1}(u_1), \dots, [\hat{F}_d^*]^{-1}(u_d)\right)$$

for some $\hat{F}^* \in \mathcal{F}$, and $\hat{F}_k^* \in \mathcal{F}_k$ for every $k = 1, \dots, d$. Note that the choice of all the parametric families is absolutely free of constraints, and that these families are not related to each other (they can be arbitrary and independently chosen). This is the usual way of generating new copula families. The price to be paid is that the true joint law F does not belong to $\mathcal{F}^*$ generally speaking. Similarly, the true marginal laws F_k do not belong to the sets $\mathcal{F}_k^*$ in general.

If a parametric assumption is made in such a case, the standard estimation procedure is semi-parametric: the copula is a function of some parameter $\theta = (\tau, \theta_1, \dots, \theta_d)$. Recall that the copula density c

is the derivative of C with respect to each of its arguments:

$$c_\theta(\boldsymbol{u}) = \frac{\partial^d}{\partial_1 \ldots \partial_d} C(\boldsymbol{u})$$

Here, the copula density c_θ itself can be calculated under a full parametric assumption. Thus, we get an estimator of θ by maximizing the log-likelihood

$$\sum_{i=1}^{T} \log c_\theta(\widehat{F}_1(X_{1i}), \ldots, \widehat{F}_d(X_{di}))$$

for some $\sqrt{T}$-convergent estimates $\widehat{F}_k(X_{ki})$ of the marginal CDFs. Obviously, we may choose $\widehat{F}_k = F_k^{(1)}$ or $F_k^{(2)}$.

Note that such an estimator is called an "omnibus estimator", and it can be seen as a maximum-likelihood estimator of θ after replacing the unobservable ranks $F_k(X_{ki})$ by the pseudo-observations. The asymptotic distribution of the estimator has been studied in Genest *et al* (1995) and Shi and Louis (1995). The main aim of semi-parametric estimation is to avoid possible misspecification of marginal distributions, which may overestimate the degree of dependence in the data (eg, see Silvapulle *et al* (2004)). Note finally that Chen and Fan (2004a, 2004b) have developed the theory of this semi-parametric estimator in a time-series context.

Thus, depending on the degree of assumptions about the joint and marginal models, there exists a wide range of possibilities for estimating copula functions as provided by Equation (2.1). The only trap to avoid is to be sure that the assumptions made for margins are consistent with those drawn for the joint law. The statistical properties of all these estimators are the usual ones, namely consistency and asymptotic normality.

THE ESTIMATION OF COPULA DENSITIES

After the estimation of C by $\hat{C}$ as in Equation (2.1), it is tempting to define an estimate of the copula density c at every $\boldsymbol{u} \in [0,1]^d$ by

$$\hat{c}(\boldsymbol{u}) = \frac{\partial^d}{\partial_1 \cdots \partial_d} \hat{C}(\boldsymbol{u})$$

Unfortunately, this works only when $\hat{C}$ is differentiable. Most of the time, this is the case when the marginal and joint CDFs are parametric or nonparametrically smoothed (by the kernel method, for instance). In the latter case and when d is "large" (more than 3), the estimation of c can be relatively poor because of the curse of dimensionality.

Nonparametric estimation procedures for the density of a copula function have already been proposed by Behnen *et al* (1985) and Gijbels and Mielniczuk (1990). These procedures rely on symmetric kernels, and have been detailed in the context of uncensored data. Unfortunately, such techniques are not consistent on the boundaries of $[0, 1]^d$. They suffer from the so-called boundary bias. Such bias can be significant in the neighbourhood of the boundaries too, depending on the size of the bandwidth. Hereafter, we will propose some solutions to cope with such issues. To ease notation and without a lack of generality, we will restrict ourselves to the bivariate case ($d = 2$). Thus, our random vector will be denoted by (X, Y) instead of (X_1, X_2).

In the following sections, we will study some properties of some kernel-based estimators, and illustrate some of these by simulations. The benchmark will be a simulated sample, whose size is $T = 1{,}000$ and that will be generated by a Frank copula with copula density

$$c^{Fr}(u, v, \theta) = \frac{\theta[1 - e^{-\theta}]e^{-\theta(u+v)}}{([1 - e^{-\theta}] - (1 - e^{-\theta u})(1 - e^{-\theta v}))^2}$$

and Kendall's tau equal to 0.5. Hence, the copula parameter is $\theta = 5.74$. This density can be seen in Figure 2.2 together with its contour plot on the right.

Nonparametric density estimation for distributions with finite support

An initial approach relies on a kernel-based estimation of the density based on the pseudo-observations $(F_{X,T}(X_i), F_{Y,T}(Y_i))$, where $F_{X,T}$ and $F_{Y,T}$ are the empirical distribution functions

$$F_{X,T}(x) = \frac{1}{T+1}\sum_{i=1}^{T} \mathbb{1}(X_i \leq x) \quad \text{and} \quad F_{Y,T}(y) = \frac{1}{T+1}\sum_{i=1}^{T} \mathbb{1}(Y_i \leq y)$$

where the factor $T + 1$ (instead of standard T, as in Deheuvels (1979) for instance) allows the avoidance of boundary problems: the quantities $F_{X,T}(X_i)$ and $F_{Y,T}(Y_i)$ are the ranks of the X_i's and the Y_i's divided by $T+1$, and therefore take values

$$\left\{\frac{1}{T+1}, \frac{2}{T+1}, \dots, \frac{T}{T+1}\right\}$$

Standard kernel-based estimators of the density of pseudo-observations yield, using diagonal bandwidth (see Wand and Jones (1995))

$$\widehat{c}_h(u,v) = \frac{1}{Th^2}\sum_{i=1}^{T} K\left(\frac{u - F_{X,T}(X_i)}{h}, \frac{v - F_{Y,T}(Y_i)}{h}\right)$$

for a bivariate kernel $K : \mathbb{R}^2 \longrightarrow \mathbb{R}$, $\int K = 1$.

The variance of the estimator can be derived, and is $\mathcal{O}((Th^2)^{-1})$. Moreover, it is asymptotically normal at every point $(u,v) \in (0,1)$:

$$\frac{\widehat{c}_h(u,v) - E(\widehat{c}_h(u,v))}{\sqrt{\operatorname{Var}(\widehat{c}_h(u,v))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

As a benchmark, Figure 2.2 shows the theoretical density of a Frank copula. In Figure 2.3 we plot the standard Gaussian kernel estimator based on the sample of pseudo-observations $(\widehat{U}_i, \widehat{V}_i) \equiv (F_{X,T}(X_i), F_{Y,T}(Y_i))$.

Recall that even if kernel estimates are consistent for distributions with unbounded support and the support is bounded, the boundary bias can yield some "ill" underestimation (even if the distribution is twice differentiable in the interior of its support).

We can explain this phenomenon easily in the univariate case. Consider a T sample $X_1, \dots, X_T$ of a positive random variable with density f. The support of their density is then $\mathbb{R}^+$. Let K denote a symmetric kernel, whose support is $[-1,+1]$. Then, for all $x \geq 0$, using a Taylor expansion, we get

$$\begin{aligned}
E(\widehat{f}_h(x)) &= \int_{-1}^{x/h} K(y) f(x - hy)\, dy \\
&= f(x) \cdot \int_{-1}^{x/h} K(y)\, dy \\
&\quad - h \cdot f'(x) \cdot \int_{-1}^{x/h} yK(y)\, dy + \mathcal{O}(h^2)
\end{aligned}$$

Figure 2.2 Density of the Frank copula with a Kendall tau equal to 0.5.

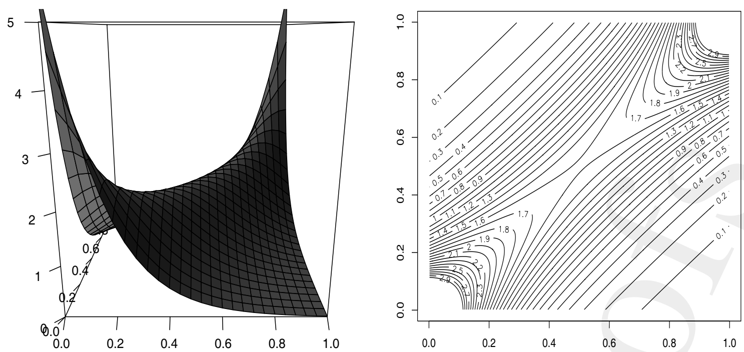
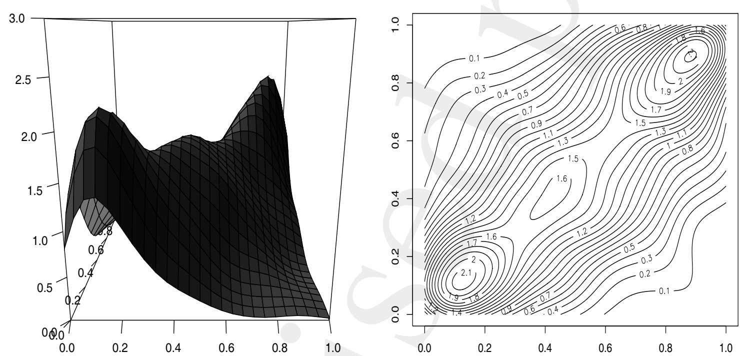


Figure 2.3 Estimation of the copula density using a Gaussian kernel based on 1,000 observations drawn from a Frank copula.

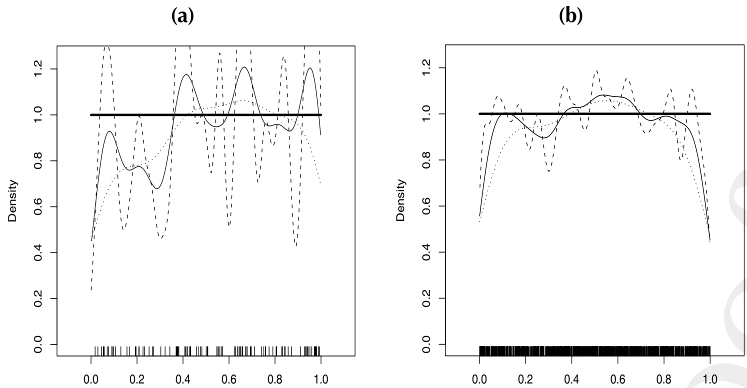


Hence, since the kernel is symmetric, $\int_{-1}^{x/h} K(y)\, dy \xrightarrow{h\to 0} 1/2$ when $x = 0$, and therefore

$$E(\widehat{f}_h(0)) = \tfrac{1}{2} f(0) + \mathcal{O}(h)$$

Note that, if $x > 0$, the expression $\int_{-1}^{x/h} K(y)\, dy$ is 1 when h is sufficiently small (when $x > h$ to be specific). Thus, this integral cannot be one, uniformly, with respect to every $x \in (0, 1]$. And for

Figure 2.4 Estimation of the uniform density on $[0, 1]$ using a Gaussian kernel and different bandwidth with (a) $n = 100$ and (b) $n = 1{,}000$ observations.



more general kernels, it has no reason to be equal to 1. In the latter case, since this expression can be calculated, normalising $\widehat{f}_h(x)$ by dividing by $\int_{-1}^{x/h} K(z)\, dz$ (at each x) achieves consistency. Nonetheless, it remains a bias that is of the order of $\mathcal{O}(h)$. Using some boundary kernels (see Gasser and Müller (1979)), it is possible to achieve $\mathcal{O}(h^2)$ everywhere in the interior of the support.

Consider the case of variables uniformly distributed on $[0, 1]$, $U_1, \ldots, U_n$. Figure 2.4 shows kernel-based estimators of the uniform density, with Gaussian kernel and different bandwidths, with $n = 100$ and 1,000 simulated variables. In that case, for any $h > 0$,

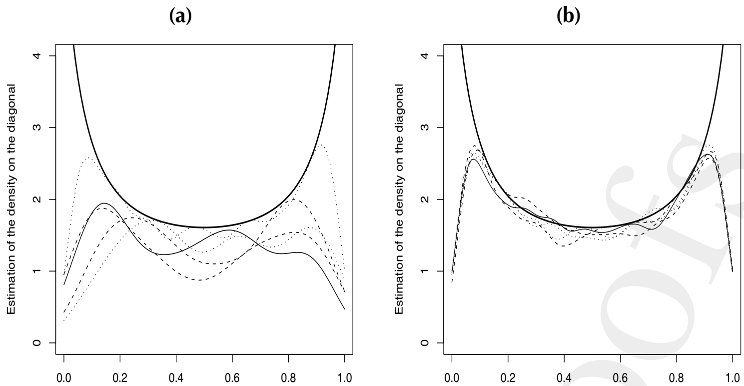
$$E(\widehat{f}_h(0)) = \int_0^1 K_h(y)\, dy = \frac{1}{h\sqrt{2\pi}} \int_0^1 \exp\left(-\frac{y^2}{2h^2}\right) dy \xrightarrow{h \to 0} \frac{1}{2} = \frac{f(0)}{2}$$

and in the interior, ie, $x \in (0, 1)$,

$$\begin{aligned} E(\widehat{f}_h(x)) &= \int_0^1 K_h(y - x)\, dy \\ &= \frac{1}{h\sqrt{2\pi}} \int_0^1 \exp\left(-\frac{(y-x)^2}{2h^2}\right) dy \xrightarrow{h \to 0} 1 = f(x) \end{aligned}$$

Dealing with bivariate copula densities, we observe the same phenomenon. On boundaries, we obtain some "multiplicative

Figure 2.5 Estimation of the copula density on the diagonal using a (standard) Gaussian kernel with (a) 100 and (b) 1,0000 observations drawn from a Frank copula.



bias", $1/4$ in corners and $1/2$ in the interior of borders. The additional bias is of the order of $\mathcal{O}(h)$ on the frontier, and standard $\mathcal{O}(h^2)$ in the interior. More precisely, in any corners (eg, $(0, 0)$)

$$E(\widehat{c}_h(0,0)) = \tfrac{1}{4}c(u,v) + \mathcal{O}(h)$$

on the interior of the borders (eg, $u = 0$ and $v \in (0, 1)$)

$$E(\widehat{c}_h(0,v)) = \tfrac{1}{2}c(u,v) + \mathcal{O}(h)$$

and in the interior ($(u, v) \in (0, 1) \times (0, 1)$)

$$E(\widehat{c}_h(u,v)) = c(u,v) + \mathcal{O}(h^2)$$

The bias can be observed in Figure 2.5, which represents the diagonal of the estimated density for several samples.

Several techniques have been introduced to obtain a better estimation on the borders for univariate densities:

- mirror image modification (Schuster (1985); Deheuvels and Hominal (1979)), where artificial data are obtained, using symmetric (mirror) transformations on the borders;
- transformed kernels (Devroye and Györfi (1985); Wand *et al* (1991)), where the idea is to transform the data X_i using a

bijective mapping ϕ so that the $\phi(X_i)$ have support $\mathbb{R}$. Efficient kernel-based estimation of the density of the $\phi(X_i)$ can be derived, and, by the inverse transformation, we get back the density estimation of the X_i themselves;

- boundary kernels (Gasser and Müller (1979); Rice (1984); Müller (1991)), where a smooth distortion is considered near the border, so that the bandwidth and the kernel shape can be modified (the closer to the border, the smaller).

Finally, the last section will briefly mention the impact of pseudo-observations, ie, working on samples

$$\{(F_{X,T}(X_1), F_{Y,T}(Y_1)), \ldots, (F_{X,T}(X_T), F_{Y,T}(Y_T))\}$$

instead of

$$\{(F_X(X_1), F_Y(Y_1)), \ldots, (F_X(X_T), F_Y(Y_T))\}$$

as if we know the true marginal distributions.

Mirror image

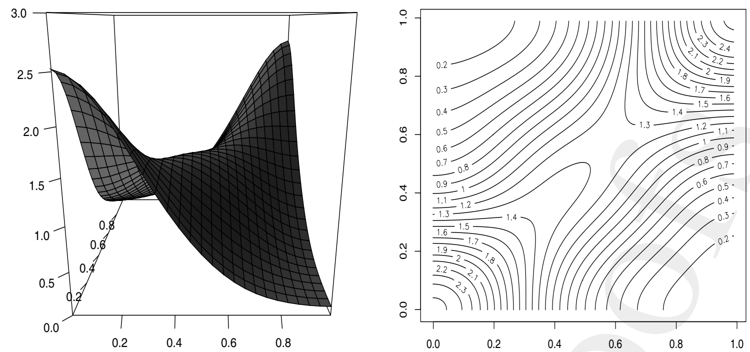
The idea of this method, developed by Deheuvels and Hominal (1979) and Schuster (1985), is to add some "missing mass" by reflecting the sample with respect to the boundaries. They focus on the case where variables are positive, ie, whose support is $[0, \infty)$. Formally and in its simplest form, it means replacing $K_h(x - X_i)$ by $K_h(x - X_i) + K_h(x + X_i)$. The estimator of the density is then

$$\widehat{f}_h(x) = \frac{1}{Th} \sum_{i=1}^{T} \left\{ K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right\}$$

In the case of densities whose support is $[0, 1] \times [0, 1]$, the non-consistency can be corrected on the boundaries, but the convergence rate of the bias will remain $\mathcal{O}(h)$ on the boundaries, which is larger than the usual rate $\mathcal{O}(h^2)$ obtained in the interior if $h \to 0$. The only case where the usual rate of convergence is obtained on boundaries is when the derivative of the density is zero on such subsets. Note that the variance is 4 times higher in corners and 2 times higher in the interior of borders.

For copulas, instead of using only the "pseudo-observations" $(\widehat{U}_i, \widehat{V}_i) \equiv (F_{X,T}(X_i), F_{Y,T}(Y_i))$, the mirror image consists in reflecting each data point with respect to all edges and corners of the

Figure 2.6 Estimation of the copula density using a Gaussian kernel and the mirror reflection principle with 1,000 observations from the Frank copula.



unit square $[0,1]\times[0,1]$. Hence, additional observations can be considered; ie, the $(\pm\widehat{U}_i, \pm\widehat{V}_i)$, the $(\pm\widehat{U}_i, 2-\widehat{V}_i)$, the $(2-\widehat{U}_i, \pm\widehat{V}_i)$ and the $(2-\widehat{U}_i, 2-\widehat{V}_i)$, so that one can consider

$$
\begin{aligned}
&\hat{c}_h(u,v)\\
&\quad=\frac{1}{Th^2}\sum_{i=1}^{T}\biggl\{K\left(\frac{u-\widehat{U}_i}{h}\right)K\left(\frac{v-\widehat{V}_i}{h}\right)+K\left(\frac{u+\widehat{U}_i}{h}\right)K\left(\frac{v-\widehat{V}_i}{h}\right)\\
&\qquad+K\left(\frac{u-\widehat{U}_i}{h}\right)K\left(\frac{v+\widehat{V}_i}{h}\right)+K\left(\frac{u+\widehat{U}_i}{h}\right)K\left(\frac{v+\widehat{V}_i}{h}\right)\\
&\qquad+K\left(\frac{u-\widehat{U}_i}{h}\right)K\left(\frac{v-2+\widehat{V}_i}{h}\right)+K\left(\frac{u+\widehat{U}_i}{h}\right)K\left(\frac{v-2+\widehat{V}_i}{h}\right)\\
&\qquad+K\left(\frac{u-2+\widehat{U}_i}{h}\right)K\left(\frac{v-\widehat{V}_i}{h}\right)+K\left(\frac{u-2+\widehat{U}_i}{h}\right)K\left(\frac{v+\widehat{V}_i}{h}\right)\\
&\qquad+K\left(\frac{u-2+\widehat{U}_i}{h}\right)K\left(\frac{v-2+\widehat{V}_i}{h}\right)\biggr\}
\end{aligned}
$$

Figure 2.6 has been obtained using the reflection principle. We can check that the fit is far better than in Figure 2.3.

Transformed kernels

Recall that c is the density of (U,V), $U=F_X(X)$ and $V=F_Y(Y)$. The two latter random variables (RVs) follow uniform distributions (marginally). Consider a distribution function G of a continuous distribution on $\mathbb{R}$, with differentiable strictly positive density g. We build new RVs $\tilde{X} = G^{-1}(U)$ and $\tilde{Y} = G^{-1}(V)$. Then, the density of $(\tilde{X}, \tilde{Y})$ is

$$f(x, y) = g(x)g(y)c[G(x), G(y)] \tag{2.2}$$

This density is twice continuously differentiable on $\mathbb{R}^2$, and the standard kernel approach applies.

Since we do not observe a sample of (U, V) but instead make pseudo-observations $(\hat{U}_i, \hat{V}_i)$, we build an "approximated sample" of the transformed variables $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_T, \tilde{Y}_T)$ by setting $\tilde{X}_i = G^{-1}(\hat{U}_i)$ and $\tilde{Y}_i = G^{-1}(\hat{V}_i)$. Thus, the kernel estimator of f is

$$\hat{f}(x, y) = \frac{1}{Th^2} \sum_{i=1}^{T} K\left(\frac{x - \tilde{X}_i}{h}, \frac{y - \tilde{Y}_i}{h}\right) \tag{2.3}$$

The associated estimator of c is then deduced by inverting (2.2),

$$c(u, v) = \frac{f(G^{-1}(u), G^{-1}(v))}{g(G^{-1}(u))g(G^{-1}(v))}, \quad (u, v) \in [0, 1] \times [0, 1]$$

and therefore we get

$$\begin{aligned} \widehat{c}_h(u, v) &= \frac{1}{Th^2 g(G^{-1}(u)) \cdot g(G^{-1}(v))} \\ &\quad \times \sum_{i=1}^{T} K\left(\frac{G^{-1}(u) - G^{-1}(\hat{U}_i)}{h}, \frac{G^{-1}(v) - G^{-1}(\hat{V}_i)}{h}\right) \end{aligned}$$

Note that this approach can be extended by considering different transformations G_X and G_Y, different kernels K_X and K_Y, or different bandwidths h_X and h_Y, for the two marginal random variables.

Figure 2.7 was obtained using the transformed kernel, where K was a Gaussian kernel and G was respectively the CDF of the $\mathcal{N}(0, 1)$ distribution.

The absence of a multiplicative bias on the borders can be observed in Figure 2.8, where the diagonal of the copula density is plotted, based on several samples. The copula density estimator obtained with transformed samples has no bias, is asymptotically normal, etc. Actually, we get all the usual properties of the multivariate kernel density estimators.

Figure 2.7 Estimation of the copula density using a Gaussian kernel and Gaussian transformations with 1,000 observations drawn from the Frank copula.

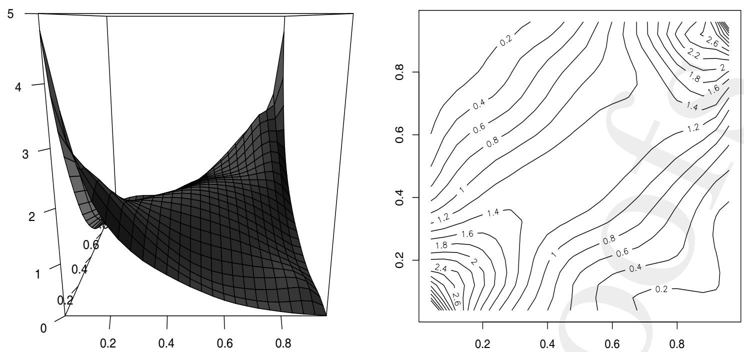
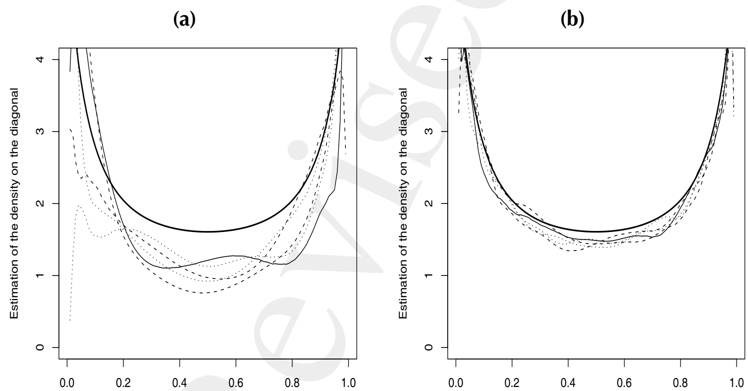


Figure 2.8 Estimation of the copula density on the diagonal using a Gaussian kernel and Gaussian transformations with (a) 100 and (b) 1,0000 observations drawn from a Frank copula.



Beta kernels

In this section we examine the use of the beta kernel introduced by Brown and Chen (1999), and Chen (1999, 2000) for nonparametric estimation of regression curves and univariate densities with compact support, respectively.

Following an idea by Harrell and Davis (1982), Chen (1999, 2000) introduced the beta kernel estimator as an estimator of a density function with known compact support $[0, 1]$, to remove the boundary bias of the standard kernel estimator:

$$\widehat{f}_h(x) = \frac{1}{T}\sum_{i=1}^{T} K\left(X_i, \frac{x}{h} + 1, \frac{1-x}{h} + 1\right)$$

where $K(\cdot, \alpha, \beta)$ denotes the density of the beta distribution with parameters α and β,

$$K(x, \alpha, \beta) = \frac{x^{\alpha}(1-x)^{\beta}}{B(\alpha, \beta)}, \quad x \in [0, 1]$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

The main difficulty when working with this estimator is the lack of a simple "rule of thumb" for choosing the smoothing parameter h.

The beta kernel has two leading advantages. First it can match the compact support of the object to be estimated. Secondly it has a flexible form and changes the smoothness in a natural way as we move away from the boundaries. As a consequence, beta kernel estimators are naturally free of boundary bias and can produce estimates with a smaller variance. Indeed we can benefit from a larger effective sample size since we can pool more data. Monte Carlo results available in these papers show that they have better performance compared to other estimators which are free of boundary bias, such as local linear (Jones (1993)) or boundary kernel (Müller (1991)) estimators. Renault and Scaillet (2004) also report better performance compared to transformation kernel estimators (Silverman (1986)). In addition, Bouezmarni and Rolin (2001, 2003) show that the beta kernel density estimator is consistent even if the true density is unbounded at the boundaries. This feature may also arise in our situation. For example the density of a bivariate Gaussian copula is unbounded at the corners $(0,0)$ and $(1,1)$. Therefore beta kernels are appropriate candidates to build well-behaved nonparametric estimators of the density of a copula function.

The beta-kernel based estimator of the copula density at point (u,v) is obtained using product beta kernels, which yields

$$\begin{aligned}\widehat{c}_h(u,v) = \frac{1}{Th^2}\sum_{i=1}^{T} K\left(X_i, \frac{u}{h}+1, \frac{1-u}{h}+1\right)\\ \times K\left(Y_i, \frac{v}{h}+1, \frac{1-v}{h}+1\right)\end{aligned}$$

Figure 2.9 shows that the shape of the product beta kernels for different values of u and v is clearly adaptive.

For convenience, the bandwidths are here assumed to be equal, but, more generally, one can consider one bandwidth per component. See Figure 2.10 for an example of an estimation based on beta kernels and a bandwidth $h = 0.05$.

Let $(u,v) \in [0,1] \times [0,1]$. The bias of $\widehat{c}(u,v)$ is of the order of h, $\widehat{c}_h(u,v) = c(u,v) + \mathcal{O}(h)$. The absence of a multiplicative bias on the boundaries can be observed on Figure 2.11, where the diagonal of the copula density is plotted, based on several samples.

On the other hand, note that the variance depends on the location. More precisely, $\text{Var}(\widehat{c}_h(u,v))$ is $\mathcal{O}((Th^{\kappa})^{-1})$, where $\kappa = 2$ in corners, $\kappa = 3/2$ in borders, and $\kappa = 1$ in the interior of $[0,1] \times [0,1]$. Moreover, as well as "standard" kernel estimates, $\widehat{c}_h(u,v)$ is asymptotically normally distributed:

$$\begin{gathered}\sqrt{Th^{\kappa'}}\,[\widehat{c}_h(u,v) - c(u,v)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma(u,v)^2)\\ \text{as } Th^{\kappa'} \to \infty \text{ and } h \to 0\end{gathered}$$

where κ' depends on the location, and where $\sigma(u,v)^2$ is proportional to $c(u,v)$.

Working with pseudo-observations

As we know, most of the time the marginal distributions of random vectors are unknown, as recalled in the first section. Hence, the associated copula density should be estimated not on samples $(F_X(X_i), F_Y(Y_i))$ but on pseudo-samples $(F_{X,T}(X_i), F_{Y,T}(Y_i))$.

Figure 2.9 Shape of bivariate beta kernels for different values of u and v. (a) $u = 0.0$, $v = 0.0$; (b) $u = 0.2$, $v = 0.0$; (c) $u = 0.5$, $v = 0.0$; (d) $u = 0.0$, $v = 0.2$; (e) $u = 0.2$, $v = 0.2$; (f) $u = 0.5$, $v = 0.2$; (g) $u = 0.0$, $v = 0.5$; (h) $u = 0.2$, $v = 0.5$; (i) $u = 0.5$, $v = 0.5$.

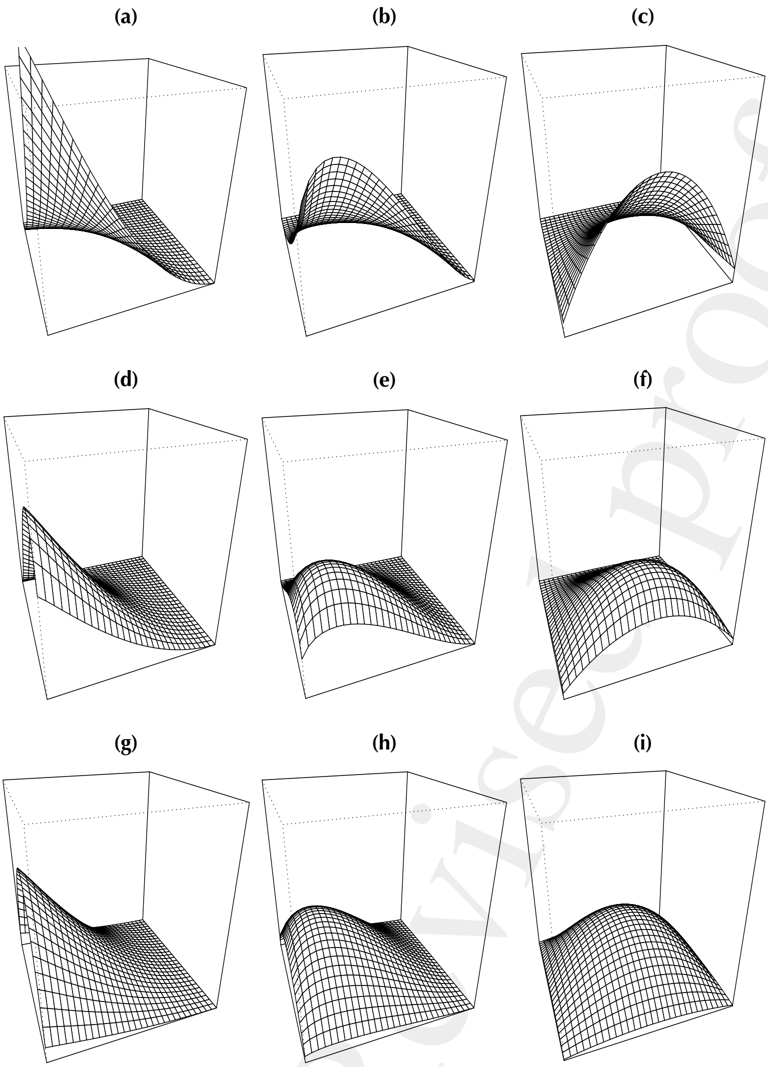


Figure 2.10 Estimation of the copula density using beta kernels ($u = 0.05$) with 1,000 observations drawn from a Frank copula.

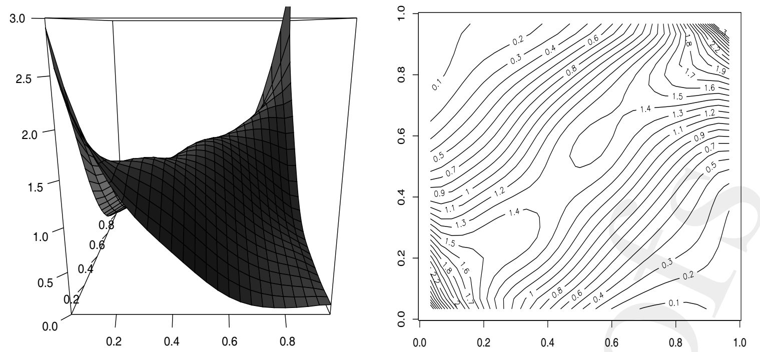


Figure 2.11 Estimation of the copula density on the diagonal using beta kernels with (a) 100 and (b) 1,000 observations drawn from a Frank copula.

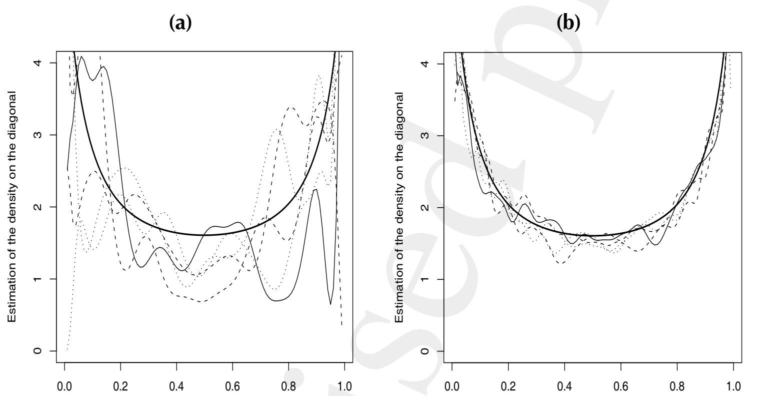
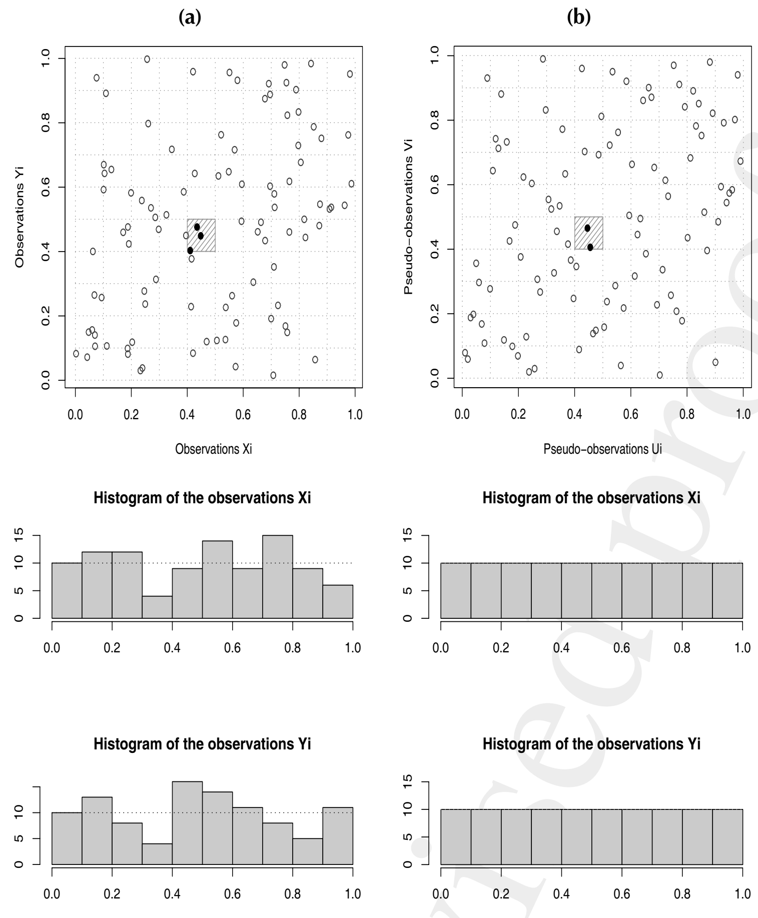


Figure 2.12 shows some scatterplots when the margins are known (ie, we know $(F_X(X_i), F_Y(Y_i))$), and when margins are estimated (ie, $(F_{X,T}(X_i), F_{Y,T}(Y_i))$). Note that the pseudo-sample is more "uniform", in the sense of a lower discrepancy (as in quasi Monte Carlo techniques; eg, see Niederreiter (1992)). Here, by mapping every point of the sample on the marginal axis, we get uniform

Figure 2.12 Scatterplots of observations and pseudo-observations and histograms of margins. (a) 100 observations (X_i, Y_i) drawn from a Frank copula and (b) associated pseudo-sample $(U_i, V_i) = (\hat{F}_X(X_i), \hat{F}_Y(Y_i))$.



grids, which is a type of "Latin hypercube" property (eg, see Jäckel (2002)).

Because samples are more "uniform" using ranks and pseudo-observations, the variance of the estimator of the density, at some given point $(u, v) \in (0, 1) \times (0, 1)$, is usually smaller. For instance,

Figure 2.13 The impact of estimating from pseudo-observations ($n = 100$). The dashed line is the distribution of $\hat{c}(u, v)$ from sample $(F_X(X_i), F_Y(Y_i))$, and the solid line is from pseudo-sample $(F_{X,T}(X_i), F_{Y,T}(Y_i))$.

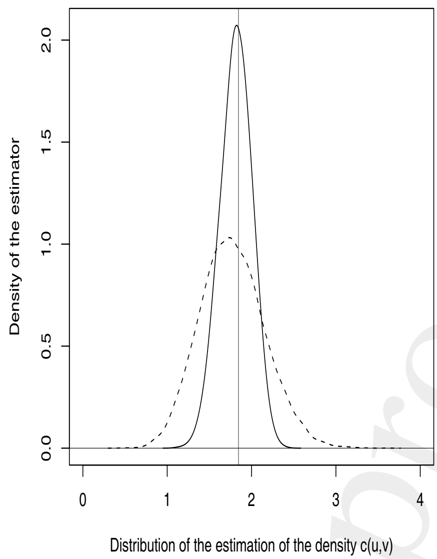


Figure 2.13 shows the impact of considering pseudo-observations, ie, substituting $F_{X,T}$ and $F_{Y,T}$ into unknown marginal distributions F_X and F_Y. The dashed line shows the density of $\hat{c}(u, v)$ from 100 observations (U_i, V_i) (drawn from the same Frank copula), and the solid line shows the density of $\hat{c}(u, v)$ from the sample of pseudo-observations (ie, the ranks of the observations).

A heuristic interpretation can be obtained from Figure 2.12. Consider the standard kernel-based estimator of the density, with a rectangular kernel. Consider a point (u, v) in the interior, and a bandwidth h such that the square $[u - h, u + h] \times [v - h, v + h]$ lies in the interior of the unit square. Given a T sample, an estimation of the density at point (u, v) involves the number of points located in the small square around (u, v). Such a number will be denoted by N, and it is a random variable. Larger N provides more precise estimations.

Assume that the margins are known, or equivalently, let $(U_1, V_1), \ldots, (U_T, V_T)$ denote a sample with distribution function C. The number of points in the small square, say N_1, is random and follows a binomial law with size T and some parameter p_1. Thus, we have $N_1 \sim \mathcal{B}(T, p_1)$ with

$$\begin{aligned} p_1 &= P((U,V) \in [u-h, u+h] \times [v-h, v+h]) \\ &= C(u+h, v+h) + C(u-h, v-h) \\ &\quad - C(u-h, v+h) - C(u+h, v-h) \end{aligned}$$

and therefore

$$\operatorname{Var}(N_1) = Tp_1(1-p_1)$$

On the other hand, assume that margins are unknown, or equivalently that we are dealing with a sample of pseudo-observations $(\widehat{U}_1, \widehat{V}_1), \ldots, (\widehat{U}_T, \widehat{V}_T)$. By construction of pseudo-observations, we have

$$\#\{\widehat{U}_i \in [u-h, u+h]\} = \lfloor 2hT \rfloor$$

where $\lfloor \cdot \rfloor$ denotes the integer part. As previously, the number of points in the small square N_2 satisfies $N_2 \sim \mathcal{B}(\lfloor 2hT \rfloor, p_2)$ where

$$\begin{aligned} p_2 &= P((\widehat{U}, \widehat{V}) \in [u-h, u+h] \times [v-h, v+h] \mid \widehat{U} \in [u-h, u+h]) \\ &= \frac{P((\widehat{U}, \widehat{V}) \in [u-h, u+h] \times [v-h, v+h])}{P(\widehat{U} \in [u-h, u+h])} \\ &\approx \big\{C(u+h, v+h) + C(u-h, v-h) \\ &\qquad - C(u-h, v+h) - C(u+h, v-h)\big\}/2h \\ &= \frac{p_1}{2h} \end{aligned}$$

Therefore the expected number of observations is the same for both methods ($E[N_1] \simeq E[N_2] \simeq Tp_1$), but

$$\operatorname{Var}(N_2) \approx 2hTp_2(1-p_2) = 2hT\frac{p_1}{2h}(1 - \frac{p_1}{2h}) = \frac{T}{2h}p_1(2h - p_1)$$

Thus

$$\frac{\operatorname{Var}(N_2)}{\operatorname{Var}(N_1)} = \frac{Tp_1(2h-p_1)}{2hTp_1(1-p_1)} = \frac{2h-p_1}{2h-2hp_1} \leq 1$$

since $h \leq 1/2$ and thus $2hp_1 \leq p_1$.

So finally, the variance of the number of observations in the small square around (u, v) is larger than the variance of the number of pseudo-observations in the same square. Therefore, this larger uncertainty concerning the relevant sub-sample used in the neighbourhood of (u, v) in the former case implies a loss of efficiency. The consequence of this result is largely counterintuitive. By working with pseudo-observations instead of "true" ones, we would expect an additional noise, which should induce more noisy estimated copula densities. This is not in fact the case as we have just shown.

CONCLUDING REMARKS

We have discussed how various estimation procedures impact the estimation of tail probabilities in a copula framework. Parametric estimation may lead to severe underestimation when the parametric model of the margins and/or the copula is misspecified. Nonparametric estimation may also lead to severe underestimation when the smoothing method does not take into account potential boundary biases in the corner of the density support. Since the primary focus of most risk management procedures is to gauge these tail probabilities, we think that the methods analysed above might help to better understand the occurrence of extreme risks in stand-alone positions (single asset) or inside a portfolio (multiple assets). In particular we have shown that nonparametric methods are simple, powerful visualisation tools that enable the detection of dependencies among various risks. A clear assessment of these dependencies should help in the design of better risk measurement tools within a VAR or an expected shortfall framework.

1 The third author acknowledges financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK).

2 For example, the number of histogram grid cells increases exponentially. This effect cannot be avoided, even using other estimation methods. Under smoothness assumptions on the density, the amount of training data required for nonparametric estimators increases exponentially with the dimension (eg, see Stone (1980)).

REFERENCES

Alexander, C., 2002, *Market Models* (New York: John Wiley and Sons).

Behnen, K., M. Huskova, and G. Neuhaus, 1985, "Rank Estimators of Scores for Testing Independence", *Statistics and Decision*, **3**, pp 239–262.

Bouezmarni, T. and J. M. Rolin, 2001, "Consistency of Beta Kernel Density Function Estimator", *The Canadian Journal of Statistics*, **31**, pp 89–98.

Bouezmarni, T. and J. M. Rolin, 2003, "Bernstein Estimator for Unbounded Density Function", Université Catholique de Louvain-la-Neuve.

Brown, B. M. and S. X. Chen, 1999, "Beta-Bernstein Smoothing for Regression Curves with Compact Supports", *Scandinavian Journal of Statistics*, **26**, pp 47–59.

Chen, S. X., 1999, "A Beta Kernel Estimation for Density Functions", *Computational Statistics and Data analysis*, **31**, pp 131–135.

Chen, S. X., 2000, "Beta Kernel for Regression Curve", *Statistica Sinica*, **10**, pp 73–92.

Chen, S. X. and Y. Fan, 2004a, "Efficient Semiparametric Estimation of Copulas", Working Paper.

Chen, S. X. and Y. Fan, 2004b, "Copula-based Tests for Dynamic Models", Working Paper.

Deheuvels, P., 1979, "La fonction de dépendance empirique et ses propriétés", *Acad. Roy. Belg., Bull. C1 Sci. 5ième sér.*, **65**, pp 274–292.

Deheuvels, P., 1981a, "A Kolmogorov–Smirnov Type Test for Independence and Multivariate Samples", *Rev. Roum. Math. Pures et Appl.*, **26**(2), pp 213–226.

Deheuvels, P., 1981b, "A Nonparametric Test for Independence, *Pub. Inst. Stat. Univ. Paris*, **26**(2), pp 29–50.

Deheuvels, P. and P. Hominal, 1979, "Estimation non paramétrique de la densité compte tenu d'informations sur le support", *Revue de Statistique Appliquée*, **27**, pp 47–68.

Devroye, L. and L. Györfi, 1985, *Nonparametric Density Estimation: The L_1 View* (New York: John Wiley and Sons).

Doukhan, P., J.-D. Fermanian and G. Lang, 2004, "Copulas of a Vector-valued Stationary Weakly Dependent Process", Working Paper.

Fermanian, J.-D., D. Radulovic and M. Wegkamp, 2004, "Weak Convergence of Empirical Copula Processes", *Bernoulli*, **10**, pp 847–860.

Fermanian, J.-D. and O. Scaillet, 2003, "Nonparametric Estimation of Copulas for Time Series", *Journal of Risk*, **95**, pp 25–54.

Fermanian, J.-D. and O. Scaillet, 2005, "Some Statistical Pitfalls in Copula Modelling for Financial Applications", in Klein, E. (ed), *Capital Formation, Governance and Banking* (New York: Nova Science Publishers), pp 59–74.

Gasser, T. and H. G. Müller, 1979, "Kernel Estimation of Regression Functions", in Gasser, T. and M. Rosenblatt (eds), *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, **757** (Springer), pp 23–68.

Genest, C., K. Ghoudi and L. P. Rivest, 1995, "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions", *Biometrika*, **82**, pp 543–552.

Gijbels, I. and J. Mielniczuk, 1990, "Estimating the Density of a Copula Function", *Communications in Statistics: Theory and Methods*, **19**, pp 445–464.

Härdle, W. and O. Linton, 1994, "Applied Nonparametric Methods", in Engle, R. and D. McFadden (eds), *Handbook of Econometrics*, **IV** (Amsterdam: North-Holland).

Harrell, F. E. and C. E. Davis, 1982, "A New Distribution-free Quantile Estimator", *Biometrika*, **69**, pp 635–640.

Jäckel, P., 2002, *Monte Carlo Methods in Finance* (New York: John Wiley and Sons).

Jones, M. C., 1993, "Simple Boundary Correction for Kernel Density Estimation", *Statistics and Computing*, **3**, pp 135–146.

Li, D. X., 1999, "On Default Correlation: a Copula Function Approach", RiskMetrics Group, Working Paper.

Longin, F. and B. Solnik, 1995, "Is the Correlation in International Equity Returns Constant: 1960–1990?", *Journal of International Money and Finance*, **14**, pp 3–26.

Müller, H. G., 1991, "Smooth Optimum Kernel Estimators near Endpoints", *Biometrika*, **78**, pp 521–530.

Niederreiter, H., 1992, "Random Number Generation and Quasi-Monte Carlo Methods", *CBMS-SIAM*, **63**, pp 86–112.

Pagan, A. and A. Ullah, 1999, *Nonparametric Econometrics* (Cambridge: Cambridge University Press).

Patton, A., 2001, "Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula", University of California, San Diego, Discussion Paper, **01-09**.

Renault, O. and O. Scaillet, 2004, "On the Way to Recovery: a Nonparametric Bias Free Estimation of Recovery Rate Densities", *Journal of Banking & Finance*, **28**, pp 2915–2931.

Rice, J. A., 1984, "Boundary Modification for Kernel Regression", *Communication in Statistics*, **12**, pp 899–900.

Schuster, E., 1985, "Incorporating Support Constraints into Nonparametric Estimators of Densities", *Communications in Statistics: Theory and Methods*, **14**, pp 1123–1136.

Shih, J. H. and T. A. Louis, 1995, "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data", *Biometrics*, **55**, pp 1384–1399.

Silvapulle, P., G. Kim and M. J. Silvapulle, 2004, "Robustness of a Semiparametric Estimator of a Copula", Econometric Society 2004 Australasian Meetings, **317**.

Silverman, B., 1986, *Density Estimation for Statistics and Data Analysis* (New York: Chapman and Hall).

Sklar, A., 1959, "Fonctions de répartition à *n* dimensions et leurs marges", *Publ. Inst. Statist. Univ. Paris*, **8**, pp 229–231.

Stone, C. J., 1980, "Optimal Rates of Convergence for Nonparametric Estimators", *Annals of Probability*, **12**, pp 361–379.

Wand, M. P. and M. C. Jones, 1995, *Kernel Smoothing* (New York: Chapman and Hall).

Wand, M. P., J. S. Marron and D. Ruppert, 1991, "Transformations in Density Estimation: Rejoinder (in Theory and Methods)", *Journal of the American Statistical Association*, **86**, pp 360–361.