Modeling dependence

Wilbert C.M. Kallenberg
Department of Applied Mathematics
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract A new way of choosing a suitable copula to model dependence is introduced. Instead of relying on a given parametric family of copulas or applying the other extreme of modeling dependence in a nonparametric way, an intermediate approach is proposed, based on a sequence of parametric models containing more and more dependency aspects. In contrast to a similar way of thinking in testing theory, the method here, intended for estimating the copula, often requires a somewhat larger number of steps. One approach is based on exponential families, another on contamination families. An extensive numerical investigation is supplied on a large number of well known copulas. The method based on contamination families is recommended. A Gaussian start in this approximation looks very promising.

Keyword and phrases: copula, Legendre polynomials, exponential family, contamination family, nonlinear correlation.

2000 Mathematics Subject Classification: 62H12, 62H20, 62P05.

1 Introduction

The classical way to deal with dependence for a multivariate distribution is to assume multivariate normality and to estimate the correlation coefficients. Also outside the normal model linear correlation is often taken as a tool to measure dependence. However, capturing only linear correlation is far too restricted. Other forms of dependence are important too. In particular in finance and insurance (see e.g. Cherubini et al. (2004), Embrechts et al. (2003), Mc Neil et al. (2005)), but also in other areas like for instance hydrology (see e.g. Genest and Favre (2006)), there is last years much more attention on going beyond the linear dependence.

A second problem with linear correlation is that the marginal distributions are mixed up with the dependence. Sklar's (1959, 1996) theorem shows that for continuous multivariate distribution functions, the univariate margins and the multivariate dependence structure can be separated, and the dependence structure can be represented by a so called copula function. This copula function is the multivariate distribution function of the random vector obtained by applying on each of the components its probability integral transformation, thus giving them uniform marginals. Embrechts et al. (2003) remark: "since linear correlation is not a copula-based measure of dependence, it can often be quite misleading and should not be taken as the canonical dependence measure". For a lot of results on copulas see also Joe (1997), Nelsen (1999), Cherubini et al. (2004), Mc Neil et al. (2005).

In view of Sklar's theorem the study of multivariate dependence can be performed in two distinct steps: estimating the marginal distributions and estimating the "intrinsic" dependence structure. The first step is very well known. Here we investigate the second step. What should be done, is choosing an appropriate (family of) copula function(s).

There are many families of copulas proposed in literature, each with their own merits. One may rely on a parametric family of copulas, like the Frank copulas or the Gumbel copulas etc. Having chosen the family one (only) needs to estimate the parameter(s) of this family. However, the choice of the parametric family is not that clear. A possible way-out is to check whether a certain copula or family of copulas is suitable. Goodness-of-fit tests for testing the simple null hypothesis of a given copula, or the composite hypothesis of a parametric family of copulas are developed, see e.g. Fermanian (2005), Panchenko (2005) and references therein. But in case of rejection, it is not clear what to do. In Biau and Wegkamp (2005) the problem of finding a particular copula, given a (parametric) class of candidate copulas is attacked. They restrict attention to copulas with a bounded density. In their oracle inequality the upper bound consists of a model error term, expressing the distance between the true density and the parametric family of candidate copulas, and a second part giving the stochastic error or estimation error. We will also split up the total error in the model error and the stochastic error, see (2.7), (2.8), (2.11) and (2.12).

The advantage of a parametric family is that only one (or a few) parameters have to be estimated, thus obtaining a relatively small stochastic error. The disadvantage might be the restriction to one family and a possible gap between the true density and the chosen family. The latter can be avoided by the other extreme of a nonparametric approach. But in that case the estimation step will lead to large errors, unless we have an enormous amount of observations. Here we propose an intermediate approach. This intermediate approach consists of two steps: a modeling step and an estimation step. In the modeling step a sequence of parametric copula models is introduced, approximating the true copula more and more. In the estimation step out of this sequence of parametric models a suitable one is selected (using a model selection rule) and subsequently the parameters within the selected model are estimated. To avoid too many technicalities we concentrate in this paper on bivariate distributions. Moreover, we concentrate in this paper on the modeling step. The estimation step will be treated in a forthcoming paper. Obviously, firstly it should be made clear that the modeling step has good approximation properties. Therefore, the aim of the present paper is to investigate the approximation error in the modeling step.

To model (and afterwards estimate) the true density of the copula, a sequence of parametric models is introduced, containing more and more dependency aspects. The method has a parametric flavor, but considering higher and higher dimensions we get in the limit the true density and in this way the method is "nonparametric". A somewhat similar approach is applied successfully in testing theory, see e.g. Kallenberg and Ledwina (1999) and Janic-Wróbleska et al. (2004). However, as a rule, in testing theory heavy forms of dependence are detected easily and therefore main attention is on copulas not too far from independence. In estimation theory the whole scope of dependent copulas should be considered carefully. This makes the modeling step more difficult.

Starting point is a given (family of) copula(s). For instance, one may simply start with the uniform density on the unit square (corresponding to independence). Another prominent starting point is the family of Gaussian copulas. Other favorite starting points of families of copulas can be used as well. Dependency aspects not covered by the starting point are added by subsequent parametric steps. In this way the method automatically improves an a priori chosen parametric family. In particular, when the starting point is not too far away from the true copula, only a few steps are needed to get a sufficiently small model error.

Well-known families of parametric models for the subsequent parametric steps are so called exponential families. For properties of exponential families we refer to Barndorff-Nielsen (1978). Approximation of (univariate) densities by exponential families has been done e.g. by Barron and Sheu (1991), Yang and Barron (1998) and Castellan (2003). Contamination families are candidates as well. The advantage of exponential families over contamination families is that the density is automatically positive and integrates to 1. However, the estimation step is more

complicated. Moreover, the marginal distributions are no longer uniform distributions, implying that fitting covariances is not equivalent to fitting correlations.

In general, two random variables X and Y are independent if and only if $cov(f_1(X), f_2(Y)) = 0$ for all f_1 and f_2 ranging over a separating class of functions (see e.g. Breiman 1968, p. 165 ff.). Eubank et al. (1987) have considered a measure of association, called ϕ^2 (see also Lancaster 1969, p.91 ff). Let $U = F_X(X), V = F_Y(Y)$, where F_X and F_Y are the marginal distribution functions of X and Y, respectively, and let b_j be the j^{th} Legendre polynomial on $(0, 1)$. If $\phi^2 < \infty$, then the condition $cov(b_r(U), b_s(V)) = 0$ for all $r, s \geq 1$ implies that X and Y are independent. So, under this mild condition, the Legendre polynomials form a separating class. Therefore, both the exponential families and the contamination families are based on suitable Legendre polynomials.

The parametric steps are designed to fit $Eb_r(U)b_s(V)$. For the contamination family this is equivalent to fitting $cov(b_r(U), b_s(V))$ or the correlation coefficient $\rho(b_r(U), b_s(V))$. For instance, when $r = s = 1$ this concerns the linear correlation of the copula. Within the exponential family the maximum likelihood estimator produces the required fit of $Eb_r(U)b_s(V)$. At the same time this member of the exponential family is closest in terms of Kullback Leibler information to the true density. For the contamination family moment estimators are invoked. They are linked up with the L_2-distance. Both for the exponential families and for the contamination families it holds that the higher the dimension of the family, the better the fit and hence the smaller the model error. On the other hand, the higher the dimension, the more parameters have to be estimated and the larger the stochastic error due to the estimation part.

It turns out that finding those parameters in the exponential family that fit $Eb_r(U)b_s(V)$ is much more difficult than the corresponding step in the contamination family. In the k-parameter exponential family k (rather complicated) equations should be solved, while the contamination family gives explicit expressions for the parameters involved. Moreover, fitting $Eb_r(U)b_s(V)$ in the contamination family gives automatically a fit of the covariance and correlation coefficient. Also the reduction in model error due to taking a larger dimension has a far more easy form for the contamination model.

The paper is organized as follows. In Section 2 the exponential families and the contamination families are introduced. Properties of these families in terms of Kullback Leibler information for the exponential families and L_2-distance for the contamination families are derived. Moreover, the decomposition of the total error in model error and stochastic error is discussed. After introduction of the Legendre polynomials and explanation of their usefulness in the present context, the way to choose particular pairs is explained. Section 3 concerns the model error. A lot of well known copulas are presented. Two forms of (a)symmetry are considered: the dependence of U on V compared to that of V on U and the possible difference in the left and right tail. Furthermore, whether linear correlation is dominant or not is taken into account. In this way a large range of interesting forms of dependence is covered. To compare the approximations with the given copulas measuring devices are given, including Kullback Leibler information, L_1- and L_2-distance and (relative) errors on rectangles. The results are given in Section 4. The method is illustrated by a large number of well known copulas. It is seen that errors of 10% when using the classical Gaussian copula are reduced in a few steps to only 1%. Even errors of more than 50% are reduced to only 4%, thus yielding for instance 0.030 as approximation for the true probability 0.029, see Table 11. As the performance of the contamination model is at least as good as that of the exponential model and the contamination family is far more easy to apply, the final recommendation will be to use the contamination model. The uniform start gives after a few steps in general accurate approximations. Furthermore, it looks very promising to apply a Gaussian start, but in that case an extra parameter has to be estimated.

2 Preliminaries

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d. random vectors with continuous distribution function $F_{X,Y}$. The marginal distribution functions of X and Y are denoted by F_X and F_Y, respectively. Studying dependence we may consider the simultaneous distribution function $F_{X,Y}$ and compare it with the product of the marginal distribution functions $F_X F_Y$ (independence case). However, in this way differences in the marginal distributions are mixed up with the dependence. Since we are interested in the dependence structure of X and Y we transform them in such a way that they have the same, that is uniform, marginals. Hence, we consider the copulas (U_i, V_i) with

$$U_i = F_X(X_i), V_i = F_Y(Y_i), i = 1, ..., n.$$

In case of independence the simultaneous density of U_i and V_i equals 1 on the unit square, but due to dependence it may have another form. We denote the true density of (U_i, V_i) by f and its distribution function by F. (Sometimes the distribution function of a copula is denoted by C and its density by c, but we prefer to use the notation F for the distribution function and f for its density.) Hence, we have the following relations $F_{X,Y}(x, y) = F\left(F_X(x), F_Y(y)\right)$ and $F(u, v) = F_{X,Y}\left(F_X^{-1}(u), F_Y^{-1}(y)\right)$.

To model and estimate the true density f of (U_i, V_i), a sequence of parametric models is introduced, containing more and more dependency aspects. As starting point we take a given copula or a given parametric family of copulas, denoted as f_0. In particular we consider the independence start, given by the uniform density on the unit square and the Gaussian copulas. The densities of the Gaussian copulas are given by

$$\frac{\varphi_2\left(\Phi^{-1}(u), \Phi^{-1}(v); \rho\right)}{\varphi\left(\Phi^{-1}(u)\right)\varphi\left(\Phi^{-1}(u)\right)},$$

where Φ denotes the standard normal distribution function, φ its density and $\varphi_2(x, y; \rho)$ the density of the bivariate normal distribution with correlation coefficient ρ (and expectations 0 and variances 1). Obviously, other families of copulas, like the Frank or Gumbel copulas or whatever can be taken as well as starting point f_0.

2.1 Exponential families

So called exponential families are nice and well-known families of parametric models. Starting with the copula density f_0 the simultaneous density of U_i and V_i is then given by

$$f_k(u, v; \theta) = f_0(u, v) \exp\left\{\sum_{j=1}^{k} \theta_j h_j(u, v) - \psi_k(\theta)\right\}, \tag{2.1}$$

where $\theta = (\theta_1, ..., \theta_k)$ and ψ_k is a normalizing function, given by

$$\psi_k(\theta) = \log \int f_0(u, v) \exp\left\{\sum_{j=1}^{k} \theta_j h_j(u, v)\right\} dudv.$$

Note that f_0 may contain an unknown parameter, which should be estimated as well. This estimation problem is considered separately: we simply take the "usual" estimator in the given parametric family. Therefore, we do not insert this potential extra parameter in the notation here.

In fact, in this way $\log(f/f_0)$ is approximated by a linear combination of the functions h_j, followed by a normalizing factor ψ_k to make its integral equal to 1. Exponential families ensure

automatically that we get densities as long as θ belongs to the natural parameter space

$$\Theta = \left\{ \theta : \int \int f_0(u,v) \exp\left\{ \sum_{j=1}^{k} \theta_j h_j(u,v) \right\} dudv < \infty \right\}.$$

Obviously, when f_0 and the functions h_j are bounded on $(0,1)\times(0,1)$, then $\Theta = \mathbb{R}^k$. On $int(\Theta)$ the function ψ_k is differentiable and its derivative, $\lambda_k(\theta)$, say, is given by

$$\lambda_k(\theta) = E_\theta h = (E_\theta h_1, ..., E_\theta h_k). \tag{2.2}$$

We denote the expectation of h_1 under f by $E_f h_1$. Often $E_{f_k(\theta)} h_1$ is abbreviated to $E_\theta h_1$ and the dimension k is suppressed and is seen from the context.

When estimating θ based on observations $(U_1, V_1), ..., (U_n, V_n)$, the maximum likelihood estimator $\widehat{\theta}$ is given by the maximum likelihood equations

$$\lambda_k(\widehat{\theta}) = \overline{h},$$

where $\overline{h} = \left(\overline{h}_1, ..., \overline{h}_k\right)$ and

$$\overline{h}_j = \frac{1}{n} \sum_{i=1}^{n} h_j(U_i, V_i).$$

Let $\widetilde{\theta}$ be the functional that adds to a distribution function G on the unit square the solution θ of the equation

$$E_\theta h = \int h dG. \tag{2.3}$$

Write F_n for the empirical distribution function based on observations $(U_1, V_1), ..., (U_n, V_n)$, that is F_n gives probability mass n^{-1} to each of the points $(U_1, V_1), ..., (U_n, V_n)$. Then we have $\int h dF_n = \overline{h}$ and hence $\widetilde{\theta}(F_n) = \widehat{\theta}$. Asymptotically when $n \to \infty$, $\overline{h} \to E_f h$ and hence $\widehat{\theta} \to \widetilde{\theta}(F)$. That is, asymptotically the maximum likelihood estimator fits $E_f h$, the expectation of h under the true distribution of (U_i, V_i). We often write $\widetilde{\theta}(F)$ simply as $\widetilde{\theta}$.

Another way of looking at the maximum likelihood estimator is by considering the Kullback Leibler information

$$\begin{aligned} K(f, f_k(\theta)) &= E_f \log\left(\frac{f}{f_k(\theta)}\right) = E_f \log f - E_f \log(f_k(\theta)) \\ &= E_f \log f/f_0 - \left\{ \sum_{j=1}^{k} \theta_j E_f h_j - \psi_k(\theta) \right\} \\ &= K(f, f_0) - \left\{ \sum_{j=1}^{k} \theta_j E_f h_j - \psi_k(\theta) \right\} \\ &= K(f, f_0) - K(f_k(\theta), f_0) + \sum_{j=1}^{k} \theta_j (E_\theta h_j - E_f h_j). \end{aligned} \tag{2.4}$$

It is immediately seen from the second line in (2.4) that minimizing $K(f, f_k(\theta))$ is equivalent to maximizing $\sum_{j=1}^{k} \theta_j E_f h_j - \psi_k(\theta)$, which gives the asymptotic version of the maximum likelihood estimator. So, asymptotically the maximum likelihood estimator chooses that member $f_k(\theta)$ of the exponential family which is closest to the true density f in terms of Kullback Leibler information.

The following proposition gives a "Pythagorean" result showing that $f_k(\widetilde{\theta})$ is the projection of f into the exponential family with "base" f_0.

Proposition 2.1 *Let $\Lambda = \lambda_k(\Theta)$. Suppose that $E_f h \in int(\Lambda)$. Then there exists a unique point $\widetilde{\theta} \in int(\Theta)$ such that $K(f, f_k(\widetilde{\theta})) = \min\{K(f, f_k(\theta)) : \theta \in \Theta\}$, or, equivalently, $E_{\widetilde{\theta}} h = E_f h$. Moreover,*

$$K(f, f_0) = K(f, f_k(\widetilde{\theta})) + K(f_k(\widetilde{\theta}), f_0). \tag{2.5}$$

Proof. The first two parts follow from standard exponential family theory. The last part is directly seen from (2.4) and (2.3). ■

We see from (2.5) the gain from adding the exponential family to f_0: the model error $K(f, f_0)$ is reduced to $K(f, f_k(\widetilde{\theta}))$, a reduction equal to $K(f_k(\widetilde{\theta}), f_0)$. This shows also the extra reduction from taking a higher dimension: when going from k to $k+1$ the model error is further reduced by an amount $K(f_{k+1}(\widetilde{\theta}_{k+1}), f_0) - K(f_k(\widetilde{\theta}_k), f_0)$. To stress that $\widetilde{\theta}$ is different for dimension k and $k+1$ we have added at this place an index for the dimension (so $\widetilde{\theta}_k$ is here the whole vector and not the k^{th} component).

The next proposition gives a nice decomposition.

Proposition 2.2 *Suppose that $E_f h \in int(\Lambda)$. For each $\theta \in \Theta$ we have*

$$K(f, f_k(\theta)) = K(f, f_k(\widetilde{\theta})) + K(f_k(\widetilde{\theta}), f_k(\theta)). \tag{2.6}$$

In particular for the maximum likelihood estimator $\widehat{\theta}$ we get

$$K(f, f_k(\widehat{\theta})) = K(f, f_k(\widetilde{\theta})) + K(f_k(\widetilde{\theta}), f_k(\widehat{\theta})). \tag{2.7}$$

Proof. It is seen from the second line in (2.4) that

$$K(f, f_k(\theta)) - K(f, f_k(\widetilde{\theta})) = \psi_k(\theta) - \psi_k(\widetilde{\theta}) + \sum_{j=1}^{k} (\widetilde{\theta}_j - \theta_j) E_f h_j.$$

Direct calculation gives

$$K(f_k(\widetilde{\theta}), f_k(\theta)) = E_{\widetilde{\theta}} \log \frac{f_k\left(\widetilde{\theta}\right)}{f_k(\theta)} = \psi_k(\theta) - \psi_k(\widetilde{\theta}) + \sum_{j=1}^{k} (\widetilde{\theta}_j - \theta_j) E_{\widetilde{\theta}} h_j.$$

Using $E_{\widetilde{\theta}} h = E_f h$ the proof of (2.6) is complete. ■

Note that (2.5) follows from (2.6) by taking $\theta = 0$. Equation (2.7) can be interpreted as

$$\text{Total Error} = \text{Model Error} + \text{Stochastic Error}. \tag{2.8}$$

The model error $K(f, f_k(\widetilde{\theta}))$ expresses how good the exponential family approximates the true density f and the stochastic error $K(f_k(\widetilde{\theta}), f_k(\widehat{\theta}))$ is due to estimation. It should be remarked that there is nothing sacred to the maximum likelihood estimator in (2.7): the decomposition of the total error into the model error and stochastic error continues to hold for any estimator. But for the maximum likelihood estimator, the stochastic error vanishes as $n \to \infty$, since $\widehat{\theta} \to \widetilde{\theta}$ as $n \to \infty$.

Summarizing, the functional $\widetilde{\theta}(F)$ maximizes $\exp\left\{\sum_{j=1}^{k} \theta_j E_f h_j - \psi_k(\theta)\right\}$, minimizes $K(f, f_k(\theta))$ and is the solution of the equation $E_\theta h = E_f h$. The maximum likelihood estimator is given by $\widehat{\theta} = \widetilde{\theta}(F_n)$. The total error is split up into the model error and the stochastic error. The reduction of the model error equals $K(f_k(\widetilde{\theta}), f_0)$.

Remark 2.1 For each density $\widetilde{f}$ satisfying $E_{\widetilde{f}}h_j = E_f h_j$ for all $j = 1, ..., k$ we have

$$K(\widetilde{f}, f_0) = K(\widetilde{f}, f_k(\widetilde{\theta})) + K(f_k(\widetilde{\theta}), f_0)$$

and hence

$$K(\widetilde{f}, f_0) \geq K(f_k(\widetilde{\theta}), f_0).$$

Therefore, among all asymptotic estimators fitting the expectation of h the one based on the exponential family is the best in the sense of Kullback Leibler information w.r.t. f_0.

2.2 Contamination families

Again we start with the copula f_0. Instead of approximating $\log(f/f_0)$ we approximate $f - f_0$ by a linear combination of the functions h_j, yielding the contamination family

$$f_k(u, v; \theta) = f_0(u, v) + \sum_{j=1}^{k} \theta_j h_j(u, v) \tag{2.9}$$

with

$$\int_0^1 h_j(u, v)du = 0, \int_0^1 h_j(u, v)dv = 0,$$

implying that the marginal distribution of both U and V under $f_k(\theta)$ is the uniform distribution. We do not require that $f_k(u, v; \theta) \geq 0$ for all u, v. Thus formally $f_k(\theta)$ is not a density: it is just an approximation of the true density f. For the functions h_j we will restrict attention to a complete orthonormal system in the L_2-sense. Hence,

$$\int_0^1 \int_0^1 h_r(u, v)h_s(u, v)dudv = \delta_{rs}$$

with $\delta_{rs} = 1$ if $r = s$ and $\delta_{rs} = 0$ if $r \neq s$. We denote the L_2-norm of f by

$$\|f\|_2 = \left\{ \int_0^1 \int_0^1 f(u, v)^2 dudv \right\}^{1/2}$$

and write the inner product of f and g as

$$< f, g >= \int_0^1 \int_0^1 f(u, v)g(u, v)dudv.$$

Let $f, f_0 \in L_2$, that is $\|f\|_2 < \infty, \|f_0\|_2 < \infty$. We then have (with equality in the L_2-sense)

$$f(u, v) - f_0(u, v) = \sum_{j=1}^{\infty} < f - f_0, h_j > h_j(u, v)$$

and $\|f - f_k(\theta)\|_2^2$ is minimized by taking $\theta_j =< f - f_0, h_j >= \widetilde{\theta}_j$, say, the j^{th} Fourier coefficient of $f - f_0$ for $j = 1, ..., k$. Hence $\widetilde{\theta}_j(F) = \int_0^1 \int_0^1 h_j dF - \int_0^1 \int_0^1 h_j dF_0$. The corresponding estimator $\widehat{\theta}_j$ based on observations $(U_1, V_1), ..., (U_n, V_n)$ is given by $\widehat{\theta}_j = \widetilde{\theta}_j(F_n) = \overline{h}_j - E_{f_0}h_j$. It is easily seen that

$$E_{\widetilde{\theta}}h = E_f h$$

as in the exponential family, but here the solution can be given in explicit form.

The following proposition is the equivalent of Proposition 2.1. It shows that $f_k(\widetilde{\theta})$ is the projection of f into the contamination family with "base" f_0.

Proposition 2.3 *Let $f, f_0 \in L_2$ and $\widetilde{\theta}_j =< f - f_0, h_j >$. Then*

$$\|f - f_0\|_2^2 = \left\|f - f_k(\widetilde{\theta})\right\|_2^2 + \left\|f_k(\widetilde{\theta}) - f_0\right\|_2^2 \tag{2.10}$$

Proof. Since $f(u,v) - f_k(u,v;\widetilde{\theta}) = \sum_{j=k+1}^{\infty} < f - f_0, h_j > h_j(u,v)$ and $f_k(u,v;\widetilde{\theta}) - f_0(u,v) = \sum_{j=1}^{k} < f - f_0, h_j > h_j(u,v)$, orthonormality of the system $h_1, h_2, ...$ gives the result. ■

We see from (2.10) the gain from adding the contamination family to f_0: the model error $\|f - f_0\|_2^2$ is reduced to $\left\|f - f_k(\widetilde{\theta})\right\|_2^2$, a reduction equal to $\left\|f_k(\widetilde{\theta}) - f_0\right\|_2^2 = \sum_{j=1}^{k} \widetilde{\theta}_j^2$. This shows also the extra reduction from taking a higher dimension: when going from k to $k+1$ the model error is further reduced by an amount $\widetilde{\theta}_{k+1}^2$.

The compeer of Proposition 2.2 reads as follows.

Proposition 2.4 *For each $\theta \in \mathbb{R}^k$ we have*

$$\|f - f_k(\theta)\|_2^2 = \left\|f - f_k(\widetilde{\theta})\right\|_2^2 + \left\|f_k(\widetilde{\theta}) - f_k(\theta)\right\|_2^2.$$

In particular, for the estimator $\widehat{\theta}$ we get

$$\left\|f - f_k(\widehat{\theta})\right\|_2^2 = \left\|f - f_k(\widetilde{\theta})\right\|_2^2 + \left\|f_k(\widetilde{\theta}) - f_k(\widehat{\theta})\right\|_2^2. \tag{2.11}$$

Proof. Because $f(u,v) - f_k(u,v;\widetilde{\theta}) = \sum_{j=k+1}^{\infty} < f - f_0, h_j > h_j(u,v)$ and $f_k(u,v;\widetilde{\theta}) - f_k(u,v;\theta) = \sum_{j=1}^{k} (< f - f_0, h_j > -\theta_j) h_j(u,v)$, orthonormality of the system $h_1, h_2, ...$ gives the result. ■

Equation (2.11) can be interpreted as

$$\text{Total Error} = \text{Model Error} + \text{Stochastic Error}. \tag{2.12}$$

The model error $\left\|f - f_k(\widetilde{\theta})\right\|_2^2$ expresses how good the contamination family approximates the true density f and the stochastic error $\left\|f_k(\widetilde{\theta}) - f_k(\widehat{\theta})\right\|_2^2$ is due to estimation.

Summarizing, the functional $\widetilde{\theta}(F)$ minimizes $\|f - f_k(\theta)\|_2^2$ and $\widetilde{\theta}_j =< f - f_0, h_j >$. The estimator $\widehat{\theta}$ is given by $\widehat{\theta} = \widetilde{\theta}(F_n)$. The total error is split up into the model error and the stochastic error. The reduction of the model error equals $\left\|f_k\left(\widetilde{\theta}\right) - f_0\right\|_2^2 = \sum_{j=1}^{k} \widetilde{\theta}_j^2$.

Comparison of the two approaches leads to the following conclusions:

1. The exponential family produces a density, but the marginals are in general not uniformly distributed. The contamination family gives not necessarily a density, but its "marginals" are uniformly distributed. In practice it turns out (see the approximation results in Section 4) that the approximation becomes negative only on a very small part of the unit square and that this causes no problems in approximating "usual" probabilities.

2. Finding $\widetilde{\theta}$ and hence also estimation is hard in the exponential family, while it is easy in the contamination family.

3. The improvement by taking a higher dimension is easily calculated in contamination families and requires more effort in exponential families.

4. Exponential families are linked up with Kullback Leibler information, while contamination families give similar results in terms of the L_2-norm. In both cases the approximation $f_k(\widetilde{\theta})$ can be seen as a projection of f on the respective family and the total error is split up into the model error and the stochastic error.

5. Fitting of the covariance of $g_1(U)$ and $g_2(V)$, for general functions g_1, g_2, is in contamination families equivalent to fitting their correlation coefficients; in fact, we may even fit the expectation of $g_1(U)\, g_2(V)$; in exponential families this does not hold.

6. Both in the exponential family and in the contamination family the expectation of h is fitted, that is $E_{\widetilde{\theta}} h = E_f h$.

2.3 Legendre polynomials

An important point is how to choose the functions $(h_1, ..., h_k)$ in (2.1) or (2.9). We start with the contamination family. A natural and very useful way to describe smooth functions on the interval $(0,1)$ is to apply the orthonormal system of Legendre polynomials. This leads for a function z on $(0,1)$ to

$$z(u) = \sum_{r \geq 0} c_r b_r(u),$$

where b_r is the r^{th} Legendre polynomial on $(0,1)$ and c_r is the r^{th} Fourier coefficient, given by $c_r =< z, b_r >= \int_0^1 z(u) b_r(u) du$. The Legendre polynomials $b_0, ..., b_5$ are given by

$$\begin{aligned}
b_0(u) &= 1 \\
b_1(u) &= \sqrt{3}(2u-1) \\
b_2(u) &= \sqrt{5}(6u^2 - 6u + 1) \\
b_3(u) &= \sqrt{7}(20u^3 - 30u^2 + 12u - 1) \\
b_4(u) &= 3(70u^4 - 140u^3 + 90u^2 - 20u + 1) \\
b_5(u) &= \sqrt{11}\left(252u^5 - 630u^4 + 560u^3 - 210u^2 + 30u - 1\right).
\end{aligned}$$

For the function $f(u,v) - f_0(u,v)$ on the unit square we have for fixed v a function on $(0,1)$ and hence

$$f(u,v) - f_0(u,v) = \sum_r c_r(v) b_r(u)$$

with $c_r(v) = \int \{f(u,v) - f_0(u,v)\} b_r(u) du$. The functions $c_r(v)$ are again functions on $(0,1)$ and hence they may be described in terms of Legendre polynomials as well. This leads to

$$f(u,v) - f_0(u,v) = \sum_{r,s} c_{rs} b_r(u) b_s(v) \tag{2.13}$$

with

$$\begin{aligned}
c_{rs} &= \int\int \{f(u,v) - f_0(u,v)\} b_r(u) b_s(v) du dv = E_f b_r(U) b_s(V) - E_{f_0} b_r(U) b_s(V) \\
&= \rho(b_r(U), b_s(V); f) - \rho(b_r(U), b_s(V); f_0)
\end{aligned}$$

for copulas f and f_0,where $\rho(b_r(U), b_s(V); f)$ denotes the correlation coefficient of $b_r(U)$ and $b_s(V)$ under f. So, c_{rs} has a nice interpretation: it is just the change of the correlation coefficient of $b_r(U)$ and $b_s(V)$, when going from f_0 to f.

In view of (2.13) a natural choice is to take for $h_1, h_2, ...$ the functions $b_r(u) b_s(v), r, s = 1, 2, ...$ Note that we may start with $r, s = 1$, because $c_{0s} = c_{r0} = 0$ for all r, s.

Apart form this analytic motivation, there is another more statistical argument. When dealing with dependence the classical and still most important way is to look at linear correlation, that is to consider the linear correlation coefficient

$$\frac{cov_f\left(U,V\right)}{\sqrt{var_f(U)var_f(V)}}=\frac{E_f(UV)-\frac{1}{4}}{\frac{1}{12}},$$

or, equivalently

$$E_f\left\{b_1(U)b_1(V)\right\}.$$

This in turn is equivalent to $\rho\left(b_1(U),b_1(V);f\right)$. When taking the contamination family with $h_1(u,v)=b_1(u)b_1(v)$ we are therefore fitting the linear correlation. Since we want to go beyond the linear correlation, we also consider $b_1(u)b_2(v), b_2(u)b_1(v), b_2(u)b_2(v)$ etc. In this way more and more dependency aspects come into the picture, eventually giving the whole class of separating functions (see the Introduction).

In the exponential family in fact we are modeling $\log f/f_0$, since

$$\frac{f_k\left(u,v;\theta\right)}{f_0(u,v)}=\exp\left\{\sum_{j=1}^{k}\theta_j h_j(u,v)-\psi_k\left(\theta\right)\right\}\Leftrightarrow\log\frac{f_k\left(u,v;\theta\right)}{f_0(u,v)}=\sum_{j=1}^{k}\theta_j h_j(u,v)-\psi_k\left(\theta\right).$$

Note that f/f_0 is always nonnegative, but $\log f/f_0$ does not have this restriction. The Legendre polynomials are a natural choice here also. As in the contamination family the Fourier coefficients $c_{rs}=E_f b_r(U)b_s(V)-E_{f_0}b_r(U)b_s(V)$ are fitted, since $E_{\widetilde{\theta}}h=E_f h$, see (2.3). However, here that does not imply that the correlation coefficients are fitted as well, because the marginals under $f_k(\widetilde{\theta})$ are in general not uniformly distributed.

Remark 2.2 It should be noted that the linear correlation of X and Y is not the same as that of $U=F_X(X)$ and $V=F_Y(Y)$. In view of Sklar's theorem we study separately the forms of the marginal distributions and the dependence structure. Therefore, linear correlation in this context refers to linear correlation of U and V.

Remark 2.3 The Legendre polynomials correspond with the first, second, third . . . moment and both in the contamination family and in the exponential family the estimator $\widehat{\theta}$ is given by $E_{\widehat{\theta}}h=\overline{h}$; hence, the estimator $\widehat{\theta}$ can also be seen as a kind of moment estimator.

2.4 Choice of pairs

The approximations by the contamination or exponential families are not yet completely described. We have to determine which pairs of Legendre polynomials are involved. Denote a contamination or exponential family symbolically by the pair of degrees of the Legendre polynomials that are involved. So, for instance $[(1,1),(2,1),(3,2)]$, refers in case of the exponential family to

$$f_3\left(u,v;\theta\right)=\exp\left\{\sum_{j=1}^{3}\theta_j h_j(u,v)-\psi_3\left(\theta\right)\right\}$$

with

$$h_1(u,v)=b_1(u)b_1(v), h_2(u,v)=b_2(u)b_1(v), h_3(u,v)=b_3(u)b_2(v).$$

With the approximation we want to reduce the model error. Therefore, we choose the pairs in such a way that we get a maximal reduction of the model error. Consider the contamination

family $[(r_1,s_1),(r_2,s_2),...,(r_k,s_k)]$. For this model we have as model error, see (2.10),

$$\left\|f-f_k(\widetilde{\theta})\right\|_2^2=\|f-f_0\|_2^2-\left\|f_k\left(\widetilde{\theta}\right)-f_0\right\|_2^2$$
$$=\|f-f_0\|_2^2-\sum_{j=1}^{k}\widetilde{\theta}_j^2.$$

Therefore, the optimal k-dimensional model is the one with the k largest $\widetilde{\theta}_j^2=c_{r_js_j}^2$ among $c_{rs}=\rho\left(b_r(U),b_s(V);f\right)-\rho\left(b_r(U),b_s(V);f_0\right)$, $r,s=1,2....$ In fact, we restrict attention to maximal degree d ($d^2\geq k$), that is using $b_r(u)b_s(v)$ for $r,s=1,...,d$. Hence, the procedure is based on the deviations of the correlation coefficients of $b_r(U)$ and $b_s(V)$ under f compared to the starting point f_0 and fits the largest k among these d^2 differences of correlation coefficients where Legendre polynomials up to degree d are involved. This can be described as follows.

Step 1 Calculate c_{rs} for $r,s=1,...,d$.

Step 2 Take the k largest (in absolute value) of these d^2 deviations of f and f_0 in terms of correlation coefficients, yielding $|c_{r_1rs_1}|\geq|c_{r_2rs_2}|\geq...\geq|c_{r_krs_k}|$. (In case of equality of more pairs to $c_{r_js_j}$: order them according to the dimension with preference for lower (maximal) dimension; so when $c_{23}=c_{14}$, then c_{23} is preferred; when the highest dimensions are equal, the preference is to the lower dimension of the other component; so when $c_{32}=c_{13}$, then c_{13} is preferred; when also that is the same then the one with the lowest first component is preferred; so when $c_{31}=c_{13}$, then c_{13} is preferred.)

Step 3 The contamination model $[(r_1,s_1),(r_2,s_2),...,(r_k,s_k)]$ is the k-dimensional contamination family

$$f_k\left(u,v;\theta\right)=f_0(u,v)+\sum_{j=1}^{k}\widetilde{\theta}_jb_{r_j}(u)b_{s_j}(v)$$

and $\widetilde{\theta}=\left(\widetilde{\theta}_1,...,\widetilde{\theta}_k\right)$ is simply given by

$$\widetilde{\theta}_j=c_{r_js_j},j=1,...,k.$$

For the exponential family we can proceed in principle in a similar way, replacing L_2-distance by Kullback Leibler information. However, from a technical point of view this is far more complicated, since all $\left(\begin{array}{c}d^2\\k\end{array}\right)$ models should be taken into account. For instance, when $d=10$ and $k=5$ we have to consider 75287520 models! Therefore, we adapt here the same strategy and perform step 1 and 2 of the contamination family as well. In the third step the exponential family $[(r_1,s_1),(r_2,s_2),...,(r_k,s_k)]$ is applied with $(r_1,s_1),(r_2,s_2),...,(r_k,s_k)$ from step 1 and 2. Here $\widetilde{\theta}=\left(\widetilde{\theta}_1,...,\widetilde{\theta}_k\right)$ is determined by solving the k equations $E_{\widetilde{\theta}}b_{r_j}b_{s_j}=E_fb_{r_j}b_{s_j},j=1,...,k$ in dimension k. (Note that even with this simplification the exponential family approach takes much more effort due to solving the k equations.)

Remark 2.4 Another way to avoid the investigation of so many models in the exponential family is to use the so called diagonal approach, where we apply in dimension k simply $(1,1),(2,2),...,(k,k)$. In testing independence, see Kallenberg and Ledwina (1999) and Janic-Wróbleska et al. (2004), this approach worked very well for symmetric distributions in the sense of exchangeable copulas, that is $f(u,v)=f(v,u)$; see also Section 3. However, it turns out that large approximation errors may appear as soon as $c_{11},c_{22},...,c_{kk}$ are not among the largest k Fourier coefficients. Therefore, we do not work out this furthermore.

Remark 2.5 Still another way is to consider $[(1,1)]$ for dimension 1, $[(1,1),(1,2),(2,1),(2,2)]$ for dimension 2 etc. But then the number of parameters is growing very fast: in dimension k we get k^2 parameters to estimate. As a consequence the stochastic error will grow too fast.

3 Model error

As is seen from (2.8) and (2.12) the total error can be decomposed into the model error and the stochastic error. In this paper we discuss the model error. To do that, we consider a lot of well known copulas and compare them with the approximations by the exponential and contamination families. Moreover, we compare them with the independence case (uniform copula) and an approach based on the bivariate normal distribution, that is the Gaussian copula, which is the classical way of dealing with dependence. The aims are threefold: firstly, we want to see how good the approximation is for smooth copulas, secondly, we want to investigate the improvement when going to a higher dimension and thirdly we want to compare the three methods: uniform, Gaussian and new proposal.

3.1 Copulas

The bivariate normal distribution is determined by its linear correlation (on the (X, Y)-level). Linear correlation is often the first quantity to look at when investigating dependence. Hence, we distinguish between copulas with dominating linear correlation and those where the linear correlation is not dominating. We define the linear dominance as the relative contribution of the linear part, measured by the quantity

$$\frac{\left\{\int\int \{f(u,v)-1)\}\, b_1(u)b_1(v)dudv\right\}^2}{\int\int \{f(u,v)-1)\}^2\, dudv} = \frac{c_{11}^2}{\sum_{r,s} c_{rs}^2}. \tag{3.1}$$

For the Gaussian copula the linear correlation is indeed dominating, but other correlations are also important. To illustrate this, calculation with Matlab gives (with f being the Gaussian copula with $\rho = 0.8$) $E_f b_1(U)b_1(V) = 0.79, E_f b_2(U)b_2(V) = 0.57, E_f b_3(U)b_3(V) = 0.38$, $E_f b_4(U)b_4(V) = 0.24, E_f b_5(U)b_5(V) = 0.16$.

Remark 3.1 A remark should here be made about this kind of copulas with density tending rather fast to infinity at the boundary of the unit square. There may occur numerical problems, for instance in Matlab, when considering areas very close to the boundary. (For example, integrating the density over such extreme areas may result in negative values due to numerical problems). To avoid this, in the calculations for the Gaussian copula with $\rho = 0.8$ we therefore have restricted attention to the square $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with $\varepsilon = 0.0000252750$. This square has probability 0.9999 under f. So, it seems no problem to restrict to this square. However, analytic calculation gives $\|f\|_2^2 = (1-\rho^2)^{-1} = 2.78$, but $\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} f^2(u,v)dudv = 2.59$. So we see a substantial decrease due to an area with probability only 10^{-4}. This is not what we want and we should be careful with interpreting the L_2-norm: as a rule it is not desirable that an area with such low probability plays an important role in the decision.

Furthermore, we take symmetry into account. There are two different forms of symmetry which are of interest. The first one concerns symmetric dependence in the sense that the dependence of U on V is the same as that of V on U. We may also speak of exchangeable copulas and it simply means that $f(u,v) = f(v,u)$. Many copulas are exchangeable, but last years there is upcoming interest in situations where this form of symmetry does not hold. A practical example where exchangeability fails, is a small firm delivering goods almost exclusively to one large firm. When the large firm comes into troubles, this may have a dramatic effect on the small firm, but not the other way around.

Another form of symmetry and asymmetry concerns the possible difference in the left and right tail. Dependency of two assets may be greater during downturns than during expansions (see Cherubini et al. (2004), for examples). The second form of symmetry is expressed by $f(u,v) = f(1-u, 1-v)$.

In view of the preceding considerations we consider the copulas presented in Table 1, thus covering all kind of situations. For more information on the densities or distribution functions of these copulas, see the Appendix. In the column linear dominance the amount of dependence in terms of $\|f-1\|_2^2 = \int\int \{f(u,v)-1)\}^2 dudv$ is given as well, cf. (3.1). Obviously, the amount of dependence for Gaussian (0.8) is much larger than for Gaussian (0.2), but the linear dominance, that is the relative contribution of the linear part, is much higher for Gaussian (0.2) than for Gaussian (0.8).

Table 1 *Copulas with their properties.*

Copula	linear dominance	exchangeable	tail symmetry
Gaussian (0.2)	$\frac{0.0366}{0.0416} = 0.88$	yes	yes
Gaussian (0.8)	$\frac{0.6172}{1.5859} = 0.39$	yes	yes
Frank (0.5)	$\frac{0.0069}{0.0069} = 1.00$	yes	yes
Plackett (2)	$\frac{0.0517}{0.0540} = 0.96$	yes	yes
Ferguson	$\frac{0.0000}{0.1000} = 0.00$	yes	yes
Gumbel (1.2)	$\frac{0.0603}{0.2177} = 0.28$	yes	no
Joe-Clayton $(1, 1.8)$	$\frac{0.4197}{2.3828} = 0.18$	yes	no
Asymmetric Gumbel $(0.9, 0.2, 3)$	$\frac{0.0601}{0.6111} = 0.10$	no	no
Gonzalo-Olmo $(0.5, 1, 1)$	$\frac{0.0603}{0.0972} = 0.62$	yes	no
Gonzalo-Olmo $(0.5, 1.5, 1)$	$\frac{0.0332}{0.0702} = 0.47$	no	no

3.2 Measuring device

For the comparison we take several quantities to measure the distance. Obviously, we take the Kullback Leibler information $K(f, f_k(\widetilde{\theta}))$ and the L_2-distance $\left\|f - f_k(\widetilde{\theta})\right\|_2$. These quantities are minimized by our procedures, the Kullback Leibler information in case of exponential families, the L_2-distance when dealing with contamination families. However, we should be very careful with such measures in case of unbounded densities, see Remark 3.1. Therefore, other measures of the distance between f and its approximation $f_k(\widetilde{\theta})$ will be considered as well. Firstly, we take the L_1-distance, defined by

$$\left\|f - f_k(\widetilde{\theta})\right\|_1 = \int\int \left|f(u,v) - f_k(u,v;\widetilde{\theta})\right| dudv.$$

This measure does not suffer from the undesirable substantial influence of areas with extreme small probability. The following lemma gives some properties of the L_1- and L_2-distance.

Lemma 3.1 *Let f and g be (measurable) functions on $[0,1]\times[0,1]$. Then*
(i) $\|f-g\|_1 \le \|f-g\|_2$.
Suppose that $\|f\|_1 < \infty$ and $\|g\|_1 < \infty$ and $\int_0^1\int_0^1 f(u,v)dudv = \int_0^1\int_0^1 g(u,v)dudv$. Then
(ii) $\sup_{A\subset[0,1]\times[0,1]} \left|\iint_A f(u,v)dudv - \iint_A g(u,v)dudv\right| = \frac{1}{2}\|f-g\|_1$.

Proof. (i) Application of Schwarz's inequality yields (with 1 the function $1(u,v)\equiv 1$)

$$\|f-g\|_1 \le \|f-g\|_2 \times \|1\|_2 = \|f-g\|_2 .$$

(ii) Let $B = \{(u,v) : f(u,v) > g(u,v)\}$. Since $\iint (f-g) = 0$, we get

$$\iint_B (f-g) = \iint_{B^c} (g-f) = \frac{1}{2}\|f-g\|_1 .$$

Hence,

$$\sup_{A\subset[0,1]\times[0,1]} \left|\iint_A f(u,v)dudv - \iint_A g(u,v)dudv\right| \ge \iint_B (f-g) = \frac{1}{2}\|f-g\|_1 .$$

On the other hand, for each (measurable) set $A \subset [0,1] \times [0,1]$ we have

$$\begin{aligned}\left|\iint_A (f-g)\right| &= \left|\iint_{A\cap B} (f-g) - \iint_{A\cap B^c} (g-f)\right| \\ &\leq \max\left\{\iint_{A\cap B} (f-g)\,, \iint_{A\cap B^c} (g-f)\right\} \\ &\leq \max\left\{\iint_{B} (f-g)\,, \iint_{B^c} (g-f)\right\} \\ &= \frac{1}{2}\left\|f-g\right\|_1 .\end{aligned}$$

This completes the proof. ■

Taking in Lemma 3.1 for f the true density and for g its approximation (which is not necessarily nonnegative!), (i) implies that the L_1-distance can be made small by reducing the L_2-distance. So, for instance in the contamination model we can make $\left\|f - f_k(\widetilde{\theta})\right\|_2$ as small as we want (provided that $\|f\|_2 < \infty$) by taking k large enough. When $\left\|f - f_k(\widetilde{\theta})\right\|_2 < \varepsilon$, then automatically $\left\|f - f_k(\widetilde{\theta})\right\|_1 < \varepsilon$. The second statement in Lemma 3.1 implies that the approximation of any probability under f has an error of at most $\frac{1}{2}\left\|f-g\right\|_1$. Obviously this is the worst case. In fact, the sets we are interested in often do cover only a part of B or B^c and/or have parts in B and parts in B^c (which gives compensation!) and hence the error due to approximation for those sets is often much smaller.

In particular, we are interested in sets of the form of rectangles and therefore we compare

$$P\left(U \leq u, V \leq v\right) = F(u,v)$$

with

$$F_{\widetilde{\theta}}(u,v)$$

for several values of u, v. When $F(u,v)$ or $1 - F(u,v)$ is small it may be more interesting to look at this measure in a relative sense, defined as

$$\frac{F_{\widetilde{\theta}}(u,v)}{F(u,v)} \text{ if } F(u,v) \leq 0.5 \text{ and } \frac{1 - F_{\widetilde{\theta}}(u,v)}{1 - F(u,v)} \text{ if } F(u,v) > 0.5.$$

Similarly, we compare

$$\overline{F}(u,v) = P(U > u, V > v) \tag{3.2}$$

with

$$\overline{F}_{\widetilde{\theta}}(u,v).$$

Often we take as a yardstick the same quantity with $f_k(\widetilde{\theta})$ replaced by the Gaussian copula with fitted correlation coefficient, given by

$$\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V).$$

In this way we may show the improvement by taking into account more forms of dependence than only the linear correlation.

4 Results

In this section we present numerical results on the model error. We take $d = 10$ and investigate $k = 0, ..., 5$. Our main attention will be on the approximation of the probabilities of the involved rectangles. We consider symmetric ones with smaller and larger probability ($u = v = 0.25, 0.4$) and asymmetric ones ($u = 0.25, v = 0.5; u = 0.5, v = 0.25$). Moreover, we add the corresponding upper tail probabilities, see (3.2). Note that also for symmetric copulas it may be of interest to consider asymmetric rectangles and/or both lower- and upper tails, because the approximation is not necessarily exchangeable and/or has tail symmetry. However, often we get tail symmetry in the approximation, due to the fact that $b_r(u) = b_r(1-u)$ if r is even and $b_r(u) = -b_r(1-u)$ if r is odd and hence, $b_r(u)b_s(v) = b_r(1-u)b_s(1-v)$ if $r+s$ is even. Indeed, often we have pairs like $(1, 3)$ or $(2, 2)$ for (r, s). When symmetry holds both for the copula and for its approximation, superfluous lines in the tables are deleted.

We start with exponential families.

4.1 Exponential family

The Gaussian distribution takes a special place within all distributions, in particular for modeling dependence. We take a Gaussian copula with $\rho = 0.2$ and $\rho = 0.8$ to see the effect of having more dependence. We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.0000206052$ when $\rho = 0.2$ and $\varepsilon = 0.0000252750$ when $\rho = 0.8$. For f_0 we take the uniform distribution on the unit square. For $\rho = 0.2$ the 5 largest $|c_{rs}|$ are $c_{11} = 0.1912, c_{13} = 0.0346, c_{31} = 0.0346, c_{22} = 0.0306$ and $c_{15} = 0.0151$. For $\rho = 0.8$ we get as the 5 largest $|c_{rs}|$: $c_{11} = 0.7857, c_{22} = 0.5726, c_{33} = 0.3780, c_{44} = 0.2415$ and $c_{55} = 0.1609$. The approximation results are presented in Table 2.

Table 2 *Approximation of the Gaussian (0.2) and (0.8) copula by exponential families with a uniform start.*

ρ	$K(f, f_k(\widetilde{\theta}))$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
0.2	0.0204	0.0021	0.0016	0.0011	0.0011	0.0010	
0.8	0.5103	0.2042	0.1537	0.1460	0.1448	0.1440	
	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
0.2	0.1307	0.0430	0.0368	0.0305	0.0294	0.0298	
0.8	0.7709	0.4761	0.4201	0.4049	0.4004	0.4000	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
0.2	0.0416	0.0046	0.0035	0.0024	0.0023	0.0021	
0.8	1.5859	0.8195	0.6479	0.6083	0.5836	0.5607	
	$F_k(0.25, 0.25; \widetilde{\theta})\,/F\,(0.25, 0.25)$						$F\,(0.25, 0.25)$
0.2	0.7470	1.016	1.024	1.031	1.033	1.028	0.0837
0.8	0.3697	1.080	1.133	1.109	1.112	1.113	0.1690
	$F_k(0.4, 0.4; \widetilde{\theta})\,/F\,(0.4, 0.4)$						$F\,(0.4, 0.4)$
0.2	0.8411	1.022	1.013	1.005	1.005	1.007	0.1902
0.8	0.5317	1.026	0.975	0.980	0.980	0.979	0.3009
	$F_k(0.25, 0.5; \widetilde{\theta})\,/F\,(0.25, 0.5)$						$F\,(0.25, 0.5)$
0.2	0.8309	1.020	1.009	1.014	1.014	1.017	0.1504
0.8	0.5388	1.106	1.065	1.059	1.060	1.059	0.2319
	$F_k(0.5, 0.25; \widetilde{\theta})\,/F\,(0.5, 0.25)$						$F\,(0.5, 0.25)$
0.2	0.8309	1.020	1.025	1.014	1.014	1.010	0.1504

The conclusions from Table 2 are as follows

1. For small ρ the model error is very small and hence the approximation works very good.

2. For large ρ the approximation is rather well, but not yet very accurate.

3. There is an enormous improvement from $k = 0$ (f_0) to $k \geq 1$.

4. The reduction by taking larger k is rather small.

5. The maximal error for probabilities given by one half times the L_1-norm is often far too pessimistic. In the examples here (which are not selected for that purpose) in fact it is 5 to more than 10 times smaller than this upper bound. For instance, $F_2(0.25, 0.25; \widetilde{\theta}) - F(0.25, 0.25) = 0.002 = 0.11 \times \frac{1}{2} \left\| f - f_2(\widetilde{\theta}) \right\|_1$ for $\rho = 0.2$ and $F_2(0.25, 0.25; \widetilde{\theta}) - F(0.25, 0.25) = 0.0225 = 0.11 \times \frac{1}{2} \left\| f - f_2(\widetilde{\theta}) \right\|_1$ for $\rho = 0.8$.

6. Due to the symmetry in the distribution and the fact that we have always pairs (r, s) with $r + s$ even, the results in the upper- and lower tail are exactly the same. For $\rho = 0.2$ some small differences occur w.r.t. exchangeability, since the approximation is sometimes not exactly exchangeable. For $\rho = 0.8$ the approximation is exchangeable too.

4.2 Testing versus estimation

For testing independence the approach using exponential families with the uniform distribution as starting point was very successful. Even restricting attention to the two-dimensional models $[(1,1),(1,2)], [(1,1),(2,1)], [(1,1),(2,2)]$ has leaded to a test with high and stable power among a wide class of alternatives, see Kallenberg and Ledwina (1999) and Janic-Wróbleska et al. (2004). The first idea could be that a similar approach is successful in estimation as well. Table 2 shows that this is not quite the case. Indeed, when the correlation of the Gaussian copula is small, like $\rho = 0.2$, the method gives very nice results. But farther away from the uniform distribution, the exponential family approach starting with a uniform distribution may lead to larger errors as the L_1-norm shows. Fortunately, as is exemplified in Table 2 very clearly, the model error for probabilities is as a rule much smaller. On the other hand, it is seen that after big improvements in the first step(s) the convergence is slow.

When testing independence a start with the null hypothesis of uniformity (corresponding to independence) is very natural. Moreover, for distributions far away from uniformity (for instance in terms of linear correlation) high power is anyhow easily obtained. The focus therefore in testing is more on what is happening in the neighborhood of the null hypothesis, where the decision requires more accuracy. In estimation theory this certainly does not hold; all types of dependence should be dealt with. This makes the modeling step more difficult.

Following this type of reasoning, firstly it is seen that the estimation problem really differs from the testing problem and secondly, that starting with the uniform distribution and using the exponential family is not recommended when a heavier form of dependence is present.

4.3 Contamination family

The Gaussian copula is considered again with the same choices as in Table 2. We get the following results.

Table 3 *Approximation of the Gaussian (0.2) and (0.8) copula by contamination families with a uniform start. A $*$ denotes that the approximation is somewhere negative and hence the Kullback-Leibler information is undefined.*

ρ	$K(f, f_k(\widetilde{\theta}))$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
0.2	0.0204	0.0019	0.0015	0.0011	0.0006	0.0005	
0.8	0.5103	$*$	$*$	$*$	$*$	$*$	
	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
0.2	0.1307	0.0395	0.0309	0.0257	0.0193	0.0177	
0.8	0.7709	0.4746	0.2556	0.1807	0.1851	0.1956	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
0.2	0.0416	0.0050	0.0038	0.0026	0.0017	0.0015	
0.8	1.5859	0.9685	0.6404	0.4974	0.4390	0.4131	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
0.2	0.7470	0.988	0.992	0.996	1.012	1.007	0.0837
0.8	0.3697	0.860	1.009	1.011	1.022	1.030	0.1690
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
0.2	0.8411	1.015	1.005	0.996	0.997	0.999	0.1902
0.8	0.5317	0.983	1.005	1.025	1.033	1.034	0.3009
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
0.2	0.8309	1.010	0.997	1.000	1.000	1.004	0.1504
0.8	0.5388	1.015	1.015	1.007	1.007	1.000	0.2319
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
0.2	0.8309	1.010	1.013	1.000	1.000	0.996	0.1504

Comparison of Table 2 and 3 shows that the results for the contamination families are better than those for the exponential families. Moreover, for large ρ the approximations of the probabilities are now also very accurate. Therefore, since finding $\widetilde{\theta}$ is hard in exponential families, especially when the dimension grows, see Section 2.2, furthermore we restrict attention to the more easy contamination approach. As the Kullback Leibler is intimately related to the exponential families, we do not consider this measure anymore.

Next we consider other copulas. Obviously, when dealing with the Gaussian copula, a Gaussian start will give automatically a perfect fit. For other copulas it is of interest to consider not only a uniform start but also a Gaussian start with $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$. For the Gaussian start the case $k = 0$ is at the same time also the classical way of modeling dependence, which will be used as a yardstick for comparison.

4.3.1 Frank

We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1-10^{-4}$ under f, yielding $\varepsilon = 0.00002500006$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = 0.0830, c_{22} = 0.0041, c_{33} = 0.0002$ and $c_{13} = c_{31} = -0.0001$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals 0.0793. When taking for f_0 the Gaussian copula with $\rho = 0.0793$, the 5 largest $|c_{rs}|$ are $c_{13} = c_{31} = -0.0141, c_{11} = 0.0074$ and $c_{15} = c_{51} = -0.0063$.

Table 4 *Approximation of the Frank(0.5)-copula by contamination families with a uniform start and a Gaussian start.*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
uniform	0.0625	0.0031	0.0001	0.0000	0.0000	0.0000	
Gaussian	0.0171	0.0126	0.0078	0.0065	0.0055	0.0044	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	0.0069	0.0000	0.0000	0.0000	0.0000	0.0000	
Gaussian	0.0006	0.0004	0.0002	0.0002	0.0002	0.0001	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	0.875	0.997	1.000	1.000	1.000	1.000	0.0714
Gaussian	0.989	0.987	0.985	0.996	0.998	1.001	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	0.917	1.000	1.000	1.000	1.000	1.000	0.1744
Gaussian	0.985	0.990	0.994	1.001	1.001	1.000	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	0.915	1.000	1.000	1.000	1.000	1.000	0.1367
Gaussian	0.988	0.994	0.992	1.000	0.998	1.000	
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
uniform	0.915	1.000	1.000	1.000	1.000	1.000	0.1367
Gaussian	0.988	0.987	0.992	1.000	1.002	1.000	

The conclusions from Table 4 are as follows.

1. The approximations are very accurate.
2. The uniform approximation with $k \geq 1$ gives better results than the classical approximation with a Gaussian copula.
3. Obviously, the Gaussian start is better than the uniform start, but for $k \geq 1$ the uniform approach is better than the Gaussian approach.
4. With a uniform start there is a big improvement from $k = 0$ (f_0) to $k \geq 1$.
5. The reduction by taking larger k is rather small.
6. The maximal error for probabilities given by one half times the L_1-norm is often far too pessimistic.
7. Due to the symmetry in the distribution and the fact that we have always pairs (r, s) with $r + s$ even, the results in the upper- and lower tail are exactly the same. Some very small differences occur w.r.t. exchangeability, since the approximation is sometimes not exactly exchangeable.

4.3.2 Plackett

We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1-10^{-4}$ under f, yielding $\varepsilon = 0.0000249828$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = 0.2273, c_{22} = 0.0467, c_{33} = 0.0091, c_{44} = 0.0017$ and $c_{55} = 0.0003$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals 0.2173. When taking for f_0 the Gaussian copula with $\rho = 0.2173$, the 5 largest $|c_{rs}|$ are $c_{13} = c_{31} = -0.0374, c_{11} = 0.0196$ and $c_{15} = c_{51} = -0.0163$.

Table 5 *Approximation of the Plackett (2)-copula by contamination families with a uniform start and a Gaussian start.*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
uniform	0.1779	0.0357	0.0070	0.0013	0.0003	0.0001	
Gaussian	0.0504	0.0411	0.0282	0.0236	0.0222	0.0220	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	0.0540	0.0023	0.0001	0.0000	0.0000	0.0000	
Gaussian	0.0053	0.0039	0.0025	0.0021	0.0018	0.0016	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	0.706	0.976	1.000	1.000	1.000	1.000	0.0885
Gaussian	0.967	0.962	0.958	0.981	0.987	0.992	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	0.800	1.000	0.999	1.000	1.000	1.000	0.2000
Gaussian	0.973	0.974	0.984	1.001	1.000	0.998	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	0.797	0.996	1.000	1.000	1.000	1.000	0.1569
Gaussian	0.964	0.986	0.982	1.000	0.996	1.001	
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
Gaussian	0.964	0.970	0.982	1.000	1.004	1.001	0.1569

The conclusions from Table 5 are quite similar to those of Table 4.

4.3.3 Ferguson

We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.00002500000$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{22} = -0.2856, c_{44} = -0.0779, c_{24} = c_{42} = 0.0640$ and $c_{66} = -0.0364$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals 0. Hence, the Gaussian start is the same as the uniform start.

Table 6 *Approximation of the Ferguson-copula by contamination families with a uniform start. (The Gaussian start is identical to the uniform start.)*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
uniform	0.2618	0.1071	0.0822	0.0686	0.0492	0.0369	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	0.1000	0.0184	0.0123	0.0082	0.0041	0.0028	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	1.333	1.066	1.053	1.028	1.002	1.002	0.0469
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	1.042	1.020	1.015	1.011	1.007	1.005	0.1536
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	1.000	1.000	1.000	1.000	1.000	1.000	0.1250
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
uniform	1.000	1.000	1.000	1.000	1.000	1.000	0.1250

The conclusions from Table 6 are similar to those of Table 4, except for conclusion 3, since here there is no difference between the Gaussian and uniform start.

4.3.4 Gumbel

We do not have tail symmetry here, but the distribution is exchangeable. We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.0000228750$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = 0.2455, c_{22} = 0.1050, c_{12} = c_{21} = 0.0614$ and $c_{33} = 0.0602$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals 0.2646. When taking for f_0 the Gaussian copula with $\rho = 0.2646$, the 5 largest $|c_{rs}|$ are $c_{13} = c_{31} = 0.0700, c_{24} = c_{42} = -0.0643$ and $c_{22} = 0.0429$.

Table 7 *Approximation of the Gumbel (1.2)-copula by contamination families with a uniform start and a Gaussian start.*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
uniform	0.1872	0.0878	0.0782	0.0758	0.0702	0.0829	
Gaussian	0.0655	0.0616	0.0567	0.0563	0.0501	0.0456	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	0.2177	0.1574	0.1463	0.1426	0.1388	0.1352	
Gaussian	0.1254	0.1217	0.1179	0.1151	0.1123	0.1097	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	0.739	1.045	1.100	1.050	1.001	1.002	0.0846
Gaussian	1.076	1.026	0.977	0.973	0.969	0.995	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	0.819	1.036	1.042	1.028	1.014	1.019	0.1954
Gaussian	1.025	1.011	0.997	1.000	1.004	1.007	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	0.814	1.039	1.039	1.039	1.002	1.000	0.1536
Gaussian	1.034	1.034	0.997	1.009	1.009	1.009	
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
uniform	0.814	1.039	1.039	1.002	1.002	1.000	0.1536
Gaussian	1.034	0.997	0.997	0.997	1.009	1.009	
	$\overline{F}_k(0.75, 0.75; \widetilde{\theta}) / \overline{F}(0.75, 0.75)$						$\overline{F}(0.75, 0.75)$
uniform	0.632	0.894	0.940	0.983	1.025	1.025	0.0989
Gaussian	0.920	0.962	1.004	1.008	1.011	1.034	
	$\overline{F}_k(0.6, 0.6; \widetilde{\theta}) / \overline{F}(0.6, 0.6)$						$\overline{F}(0.6, 0.6)$
uniform	0.791	1.000	1.006	1.020	1.033	1.038	0.2023
Gaussian	0.990	1.003	1.017	1.013	1.010	1.012	
	$\overline{F}_k(0.75, 0.5; \widetilde{\theta}) / \overline{F}(0.75, 0.5)$						$\overline{F}(0.75, 0.5)$
uniform	0.776	0.991	0.991	0.991	1.025	1.023	0.1610
Gaussian	0.986	0.986	1.021	1.009	1.009	1.009	
	$\overline{F}_k(0.5, 0.75; \widetilde{\theta}) / \overline{F}(0.5, 0.75)$						$\overline{F}(0.5, 0.75)$
uniform	0.776	0.991	0.991	1.025	1.025	1.023	0.1610
Gaussian	0.986	1.021	1.021	1.021	1.009	1.009	

The conclusions from Table 7 are as follows.

1. The approximations are accurate for $k \geq 3$.
2. For the criteria, the classical approximation with a Gaussian copula gives (much) better results than the uniform approximation. For the probabilities the uniform approximation (with $k \geq 3$) often outperforms the Gaussian copula.
3. The uniform and the Gaussian approach have similar overall type of performance for the probabilities, except for $k = 0$, where the Gaussian copula is better than the uniform copula; the Gaussian approach gives better results for the criteria.

4. With a uniform start there is a big improvement from $k = 0$ (f_0) to $k \geq 1$ and substantial further improvement for larger k.

5. The reduction by taking k larger than 4 is rather small.

6. The maximal error for probabilities given by one half times the L_1-norm is often far too pessimistic.

7. We do not have tail symmetry here, but the distribution is still exchangeable. Indeed, we see some differences for the upper- and lower tail. Only for $k = 3$ (uniform start) or $k = 1, 3$ (Gaussian start) the approximation is not exchangeable and hence only in that case some differences occur.

4.3.5 Joe-Clayton

We do not have tail symmetry, but the distribution is exchangeable. We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.0000298306$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = 0.6479, c_{22} = 0.4881, c_{33} = 0.3338, c_{44} = 0.2511$ and $c_{55} = 0.2014$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals 0.6962. When taking for f_0 the Gaussian copula with $\rho = 0.6962$, the 5 largest $|c_{rs}|$ are $c_{44} = 0.1224, c_{55} = 0.1170, c_{33} = 0.1077, c_{66} = 0.1070$ and $c_{77} = 0.0976$.

Table 8 *Approximation of the Joe-Clayton-copula by contamination families with a uniform start and a Gaussian start.*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
uniform	0.5962	0.3749	0.2404	0.2547	0.2782	0.2911	
Gaussian	0.1138	0.1419	0.1534	0.1571	0.1641	0.1694	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	2.3828	1.9629	1.7245	1.6130	1.5499	1.5093	
Gaussian	0.8771	0.8621	0.8484	0.8368	0.8254	0.8158	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	0.417	0.874	1.017	1.019	1.032	1.043	0.1497
Gaussian	1.000	1.006	1.013	1.014	1.014	1.016	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	0.597	1.014	1.035	1.055	1.064	1.066	0.2681
Gaussian	1.032	1.037	1.038	1.044	1.047	1.047	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	0.592	1.023	1.023	1.015	1.015	1.006	0.2112
Gaussian	1.028	1.028	1.023	1.020	1.020	1.018	
	$\overline{F}_k(0.75, 0.75; \widetilde{\theta}) / \overline{F}(0.75, 0.75)$						$\overline{F}(0.75, 0.75)$
uniform	0.421	0.882	1.026	1.028	1.041	1.053	0.1483
Gaussian	1.009	1.016	1.022	1.023	1.023	1.025	
	$\overline{F}_k(0.6, 0.6; \widetilde{\theta}) / \overline{F}(0.6, 0.6)$						$\overline{F}(0.6, 0.6)$
uniform	0.600	1.019	1.040	1.061	1.070	1.071	0.2668
Gaussian	1.037	1.042	1.043	1.049	1.052	1.052	
	$\overline{F}_k(0.75, 0.5; \widetilde{\theta}) / \overline{F}(0.75, 0.5)$						$\overline{F}(0.75, 0.5)$
uniform	0.596	1.030	1.030	1.022	1.022	1.013	0.2097
Gaussian	1.035	1.035	1.030	1.027	1.027	1.026	

The conclusions from Table 8 are as follows.

1. For the probabilities under consideration the approximations are in general accurate for $k \geq 2$; in particular, when taking into account the rather large L_1-error, the errors in the probabilities are small.

2. The classical approximation with the Gaussian copula is often somewhat better than the approximations with the uniform start. Obviously, this is due to the large correlations for the Joe-Clayton-copula, see the values of c_{rs}, when f_0 is the uniform distribution.

3. The Gaussian approach is in general better than the uniform approach.

4. With a uniform start there is a enormous improvement from $k = 0$ (f_0) to $k \geq 1$ and sometimes a substantial further improvement for larger k. With a Gaussian start there is no substantial improvement for larger k.

5. The improvement by taking k larger than 4 (uniform start) or taking k larger than 0 (Gaussian start) is rather small. There maybe even some decline.

6. The maximal error for probabilities given by one half times the L_1-norm is far too pessimistic.

7. We do not have tail symmetry here, but the distribution is still exchangeable. Indeed, we see some differences for the upper- and lower tail. The approximations are also exchangeable and hence there are no differences w.r.t. exchangeability.

4.3.6 Asymmetric Gumbel

We do not have tail symmetry and the distribution is not exchangeable. We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.0000243734$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = 0.2452, c_{21} = 0.1849, c_{32} = 0.1414, c_{42} = 0.1393$ and $c_{31} = 0.1113$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals 0.2674. When taking for f_0 the Gaussian copula with $\rho = 0.2674$, the 5 largest $|c_{rs}|$ are $c_{21} = 0.1849, c_{32} = 0.1414, c_{42} = 0.1172, c_{63} = 0.1069$ and $c_{52} = 0.1064$.

The conclusions from Table 9 are as follows.

1. For the probabilities under consideration the approximations are in general accurate for $k \geq 2$, with somewhat larger approximation error for $F(0.25, 0.25)$ and $\overline{F}(0.75, 0.75)$; in particular, when taking into account the rather large L_1-error, the errors in the probabilities are relatively small.

2. The classical approximation with a Gausssian copula shows large errors, which are substantially improved by the approximation using the contamination family.

3. For $k = 0$ the Gaussian approach is better than the uniform approach, but for $k \geq 2$ in most cases there is no much difference between the two.

4. With a uniform start there is a enormous improvement from $k = 0$ (f_0) to $k \geq 1$ and sometimes a substantial further improvement for larger k. With a Gaussian start there is often substantial improvement for larger k.

5. The reduction by larger k is not impressive, which is seen in the slow convergence of the L_1- and L_2-norm.

6. The maximal error for probabilities given by one half times the L_1-norm is far too pessimistic.

7. We do not have tail symmetry here and the distribution is neither exchangeable. Indeed, we see some differences for the upper- and lower tail, and when changing the components.

Table 9 *Approximation of the Asymmetric Gumbel* $(0.9, 0.2, 3)$*-copula by contamination families with a uniform start and a Gaussian start.*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
uniform	0.2654	0.2431	0.1960	0.1884	0.1872	0.1636	
Gaussian	0.2382	0.1853	0.1692	0.1611	0.1682	0.1540	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	0.6111	0.5510	0.5168	0.4968	0.4774	0.4650	
Gaussian	0.5184	0.4842	0.4642	0.4505	0.4390	0.4277	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	0.761	1.076	0.923	0.912	0.881	0.894	0.0821
Gaussian	1.112	0.959	0.948	0.921	0.922	0.947	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	0.835	1.056	1.013	1.023	1.016	0.985	0.1916
Gaussian	1.047	1.004	1.014	1.008	1.014	1.011	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	0.871	1.111	0.994	0.994	0.994	1.005	0.1435
Gaussian	1.109	0.992	0.992	0.992	0.990	0.990	
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
uniform	0.771	0.984	0.984	1.015	1.015	0.978	0.1620
Gaussian	0.982	0.982	1.012	1.012	1.012	0.998	
	$\overline{F}_k(0.75, 0.75; \widetilde{\theta}) / \overline{F}(0.75, 0.75)$						$\overline{F}(0.75, 0.75)$
uniform	0.656	0.928	1.060	1.070	1.043	1.054	0.0952
Gaussian	0.959	1.091	1.100	1.078	1.077	1.055	
	$\overline{F}_k(0.6, 0.6; \widetilde{\theta}) / \overline{F}(0.6, 0.6)$						$\overline{F}(0.6, 0.6)$
uniform	0.808	1.022	1.063	1.054	1.046	1.017	0.1980
Gaussian	1.013	1.055	1.045	1.039	1.033	1.036	
	$\overline{F}_k(0.75, 0.5; \widetilde{\theta}) / \overline{F}(0.75, 0.5)$						$\overline{F}(0.75, 0.5)$
uniform	0.713	0.909	1.005	1.005	1.005	1.014	0.1753
Gaussian	0.907	1.003	1.003	1.003	1.005	1.005	
	$\overline{F}_k(0.5, 0.75; \widetilde{\theta}) / \overline{F}(0.5, 0.75)$						$\overline{F}(0.5, 0.75)$
uniform	0.850	1.084	1.084	1.051	1.051	1.010	0.1471
Gaussian	1.081	1.081	1.048	1.048	1.048	1.064	

4.3.7 Gonzalo-Olmo

Firstly, we consider the Gonzalo-Olmo (0.5,1,1)-copula. Note that we do not have tail symmetry here, but the distribution is exchangeable. We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.0000238138$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = -0.2456, c_{12} = c_{21} = -0.1070, c_{22} = 0.0726$ and $c_{23} = 0.0403$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals -0.2489. When taking for f_0 the Gaussian copula with $\rho = -0.2489$, the 5 largest $|c_{rs}|$ are $c_{12} = c_{21} = 0.1069, c_{13} = c_{31} = -0.0403$ and $c_{22} = 0.0250$.

Table 10 *Approximation of the Gonzalo-Olmo (0.5,1,1)-copula by contamination families with a uniform start and a Gaussian start.*

start	$\left\|f - f_k(\widetilde{\theta})\right\|_1$						
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
uniform	0.2203	0.1474	0.1091	0.0674	0.0601	0.0529	
Gaussian	0.1391	0.0985	0.0471	0.0465	0.0460	0.0464	
	$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$						
uniform	0.0972	0.0369	0.0254	0.0140	0.0087	0.0071	
Gaussian	0.0296	0.0182	0.0068	0.0066	0.0065	0.0059	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	2.461	1.441	1.154	0.868	0.993	1.003	0.0254
Gaussian	1.526	1.239	0.953	0.957	0.961	1.004	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	1.435	1.055	1.012	0.969	0.976	0.972	0.1115
Gaussian	1.102	1.059	1.016	1.011	1.006	1.009	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	1.535	1.111	1.111	0.992	0.992	0.975	0.0814
Gaussian	1.146	1.146	1.026	1.019	1.021	1.021	
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
uniform	1.535	1.111	0.992	0.992	0.992	0.992	0.0814
Gaussian	1.146	1.026	1.026	1.028	1.021	1.021	
	$\overline{F}_k(0.75, 0.75; \widetilde{\theta}) / \overline{F}(0.75, 0.75)$						$\overline{F}(0.75, 0.75)$
uniform	1.195	0.699	0.839	0.978	1.039	1.034	0.0523
Gaussian	0.741	0.880	1.019	1.021	1.023	1.044	
	$\overline{F}_k(0.6, 0.6; \widetilde{\theta}) / \overline{F}(0.6, 0.6)$						$\overline{F}(0.6, 0.6)$
uniform	1.225	0.900	0.936	0.973	0.979	0.984	0.1306
Gaussian	0.940	0.977	1.013	1.009	1.004	1.007	
	$\overline{F}_k(0.75, 0.5; \widetilde{\theta}) / \overline{F}(0.75, 0.5)$						$\overline{F}(0.75, 0.5)$
uniform	1.225	0.887	0.887	0.982	0.982	0.996	0.1020
Gaussian	0.914	0.914	1.009	1.004	1.005	1.005	
	$\overline{F}_k(0.5, 0.75; \widetilde{\theta}) / \overline{F}(0.5, 0.75)$						$\overline{F}(0.5, 0.75)$
uniform	1.225	0.887	0.982	0.982	0.982	0.982	0.1020
Gaussian	0.914	1.009	1.009	1.011	1.005	1.005	

The conclusions from Table 10 are as follows.

1. The approximations are accurate for $k \geq 3$, except for $F(0.25, 0.25)$ in case of a uniform start, where $k \geq 4$ is needed.

2. The classical approximation with a Gaussian copula may give large errors. The approximations (either based on the uniform or on the Gaussian copula) outperform the classical approximation tremendously, provided k is large enough.

3. The uniform and the Gaussian approach have similar overall type of performance for larger k; for small k the Gaussian copula is (slightly) better, but both approximations are not accurate enough for small k.

4. There is a big improvement from $k = 0$ (f_0) to $k \geq 1$ and substantial further improvement for larger k.

5. The reduction by taking k larger than 4 is rather small.

6. The maximal error for probabilities given by one half times the L_1-norm is often far too pessimistic.

7. We do not have tail symmetry here, but the distribution is still exchangeable. Indeed, we see some differences for the upper- and lower tail. Only for $k = 2, 5$ (uniform start) or $k = 1, 3$ (Gaussian start) the approximation is not exchangeable and hence only in that case some differences occur w.r.t. exchangeability.

Next, we consider the Gonzalo-Olmo (0.5,1.5,1)-copula. This copula does not have tail symmetry and is also not exchangeable. We truncate the unit square to $[\varepsilon, 1-\varepsilon] \times [\varepsilon, 1-\varepsilon]$ with ε such that its probability equals $1 - 10^{-4}$ under f, yielding $\varepsilon = 0.0000243891$. When taking for f_0 the uniform distribution on the unit square, the 5 largest $|c_{rs}|$ are $c_{11} = -0.1823, c_{12} = 0.1502, c_{21} = 0.0734, c_{23} = -0.0597$ and $c_{13} = -0.0355$. The fitted correlation coefficient $\rho = E_f \Phi^{-1}(U)\Phi^{-1}(V)$ equals -0.1865. When taking for f_0 the Gaussian copula with $\rho = -0.1865$, the 5 largest $|c_{rs}|$ are $c_{12} = 0.1501, c_{21} = 0.0734, c_{23} = -0.0597, c_{33} = 0.0282$ and $c_{14} = 0.0197$.

Table 11 *Approximation of the Gonzalo-Olmo (0.5,1.5,1)-copula by contamination families with a uniform start and a Gaussian start.*

start			$\left\|f - f_k(\widetilde{\theta})\right\|_1$				
uniform	0.1796	0.1501	0.0857	0.0589	0.0474	0.0389	
Gaussian	0.1463	0.0801	0.0511	0.0351	0.0302	0.0294	
			$\left\|f - f_k(\widetilde{\theta})\right\|_2^2$				
uniform	0.0702	0.0370	0.0144	0.0090	0.0055	0.0042	
Gaussian	0.0349	0.0124	0.0070	0.0034	0.0026	0.0022	
	$F_k(0.25, 0.25; \widetilde{\theta}) / F(0.25, 0.25)$						$F(0.25, 0.25)$
uniform	2.148	1.487	1.136	0.964	0.978	0.965	0.0291
Gaussian	1.528	1.177	1.005	1.018	1.019	1.039	
	$F_k(0.4, 0.4; \widetilde{\theta}) / F(0.4, 0.4)$						$F(0.4, 0.4)$
uniform	1.315	1.056	1.001	0.974	0.968	0.983	0.1216
Gaussian	1.087	1.032	1.005	0.998	1.002	1.009	
	$F_k(0.25, 0.5; \widetilde{\theta}) / F(0.25, 0.5)$						$F(0.25, 0.5)$
uniform	1.343	1.067	1.067	0.996	0.974	0.994	0.0931
Gaussian	1.088	1.088	1.017	0.994	0.993	0.993	
	$F_k(0.5, 0.25; \widetilde{\theta}) / F(0.5, 0.25)$						$F(0.5, 0.25)$
uniform	1.443	1.147	0.990	0.990	0.990	0.984	0.0866
Gaussian	1.169	1.012	1.012	1.012	1.010	1.019	
	$\overline{F}_k(0.75, 0.75; \widetilde{\theta}) / \overline{F}(0.75, 0.75)$						$\overline{F}(0.75, 0.75)$
uniform	1.088	0.753	0.931	1.018	1.012	1.005	0.0574
Gaussian	0.774	0.952	1.039	1.032	1.033	1.023	
	$\overline{F}_k(0.6, 0.6; \widetilde{\theta}) / \overline{F}(0.6, 0.6)$						$\overline{F}(0.6, 0.6)$
uniform	1.131	0.909	0.956	0.979	0.985	0.998	0.1414
Gaussian	0.935	0.982	1.005	1.011	1.014	1.009	
	$\overline{F}_k(0.75, 0.5; \widetilde{\theta}) / \overline{F}(0.75, 0.5)$						$\overline{F}(0.75, 0.5)$
uniform	1.149	0.914	0.914	0.975	0.994	1.012	0.1087
Gaussian	0.931	0.931	0.993	1.012	1.010	1.010	
	$\overline{F}_k(0.5, 0.75; \widetilde{\theta}) / \overline{F}(0.5, 0.75)$						$\overline{F}(0.5, 0.75)$
uniform	1.101	0.875	0.995	0.995	0.995	0.991	0.1135
Gaussian	0.892	1.012	1.012	1.012	1.011	1.004	

The conclusions from Table 11 are as follows.

1. For the probabilities under consideration the approximations are in general accurate for $k \geq 3$.

2. The classical approximation with a Gaussian copula may give large errors. The approximations (either based on the uniform or on the Gaussian copula) outperform the classical approximation tremendously, provided k is large enough.

3. The uniform and the Gaussian approach have similar overall type of performance for larger k; for small k the Gaussian copula is (slightly) better, but both approximations are not accurate enough for small k.

4. There is a big improvement from $k = 0$ (f_0) to $k \geq 1$ and substantial further improvement for larger k.

5. The reduction by taking k larger than 3 is rather small.

6. The maximal error for probabilities given by one half times the L_1-norm is far too pessimistic.

7. The diagonal approach (not presented here) gives worse results than the optimal approach for the considered probabilities and criteria and is not accurate enough.

8. We do not have tail symmetry here, and neither exchangeability. Indeed, we see some differences for the upper- and lower tail and some differences between (the approximations of) $F(x, y)$ and $F(y, x)$. In particular, the classical Gaussian approximation of $F(0.5, 0.25)$ shows a much larger error than the one of $F(0.25, 5)$. Note that in both cases the approximation by contamination families with $k = 5$ successfully reduces the errors of 9 and 17 % to only 1 and 2%, respectively.

4.4 Conclusions

From the numerical results obtained for the preceding copulas and other copulas, not presented here, we draw the following conclusions.

1. For the probabilities under consideration the approximations are in general accurate provided that k is large enough. In practice, parameters should be estimated and the smaller k the smaller the estimation error will be. The great scale of copulas considered here shows that often a not too large k is sufficient. This is a very good perspective for the estimation step, where k will be selected with a suitable selection rule and subsequently the parameters are estimated.

2. The classical approximation with a Gaussian copula may give in several cases large errors. The approximations (either based on the uniform or on the Gaussian copula) outperform the classical approximation in these cases tremendously, provided k is large enough. Therefore, simply using the Gaussian copula can be improved enormously by the proposed approximations. In case the Gaussian copula (or the uniform copula) already behaves rather well, the selection rule in the estimation step will indicate to stop immediately and to use the Gaussian (or uniform) copula. So, the classical Gaussian approximation will be improved in an efficient way.

3. The uniform and the Gaussian approach have in general similar overall type of performance for larger k; for small k the Gaussian copula is often (slightly) better. Therefore, an approximation with a Gaussian start can be recommended.

4. There is often a big improvement from $k = 0$ (f_0) to $k \geq 1$ and substantial further improvement for larger k.

5. The reduction by taking k larger than 3 or 4 is rather small. This is a good perspective for the selection rule: only a few steps have to be made; on the other hand, simply taking always $k = 4$ will produce unnecessary estimation error and therefore a selection step is very welcome.

6. The maximal error for probabilities given by one half times the L_1-norm is far too pessimistic.

7. The approach works well even when we do not have tail symmetry and/or exchangeability.

All in all we can conclude that the proposed modeling step has nice approximation properties.

Appendix

The densities or distribution functions of the copulas in this paper are given below.

Gaussian copula

$$f(u,v;\rho)=\frac{1}{\sqrt{1-\rho^2}}\exp\left(\frac{\zeta_1^2+\zeta_2^2}{2}+\frac{2\rho\zeta_1\zeta_2-\zeta_1^2-\zeta_2^2}{2\left(1-\rho^2\right)}\right)$$

with

$$\zeta_1=\Phi^{-1}(u),\zeta_2=\Phi^{-1}(v).$$

Frank

$$f(u,v;\alpha)=\frac{\alpha g(1;\alpha)\left\{1-g(u+v;\alpha)\right\}}{\left\{g(1;\alpha)-g(u;\alpha)g(v;\alpha)\right\}^2}$$

with

$$g(z;\alpha)=1-e^{-\alpha z}.$$

Plackett

$$f(u,v;\alpha)=\alpha\left[1+(\alpha-1)\left(u+v-2uv\right)\right]\left[\left\{1+(\alpha-1)\left(u+v\right)\right\}^2-4\alpha\left(\alpha-1\right)uv\right]^{-3/2}.$$

Ferguson

$$f(u,v)=3\left|u-v\right|\left(1-\left|u-v\right|\right)+3\left(1-\left|1-u-v\right|\right)\left|1-u-v\right|.$$

Gumbel

$$F(u,v;\alpha)=\exp\left\{-\left[(-\log u)^\alpha+(-\log v)^\alpha\right]^{1/\alpha}\right\}.$$

Joe-Clayton

$$F(u,v;\alpha,\beta)=1-\left[1-\left\{\left[1-(1-u)^\alpha\right]^{-\beta}+\left[1-(1-v)^\alpha\right]^{-\beta}-1\right\}^{-1/\beta}\right]^{1/\alpha}.$$

Asymmetric Gumbel

$$F(u,v;\alpha,\beta,\gamma)=u^{1-\alpha}v^{1-\beta}\exp\left\{-\left[(-\alpha\log u)^\gamma+(-\beta\log v)^\gamma\right]^{1/\gamma}\right\}.$$

Gonzalo-Olmo

$$F(u,v;\alpha,\beta,\gamma)=\exp\left\{-D(u,v;\alpha,\beta)\left[(-\log u)^\gamma+(-\log v)^\gamma\right]^{1/\gamma}\right\}$$

with

$$D(u,v;\alpha,\beta)=\exp\left\{\alpha\left(1-u\right)\left(1-v\right)^\beta\right\}.$$

References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, Chichester.

Barron, A. R. and Sheu, C-H. Approximation of density functions by sequences of exponential families. *Ann. Statist.* **19** 1347-1369.

Biau, G. and Wegkamp, M. (2005). A note on minimum distance estimation of copula densities. *Statist. Probab. Lett.* **73** 105-114.

Breiman, L. (1968). *Probability.* Addison-Wesley, Reading.

Castellan G. (2003). Density estimation via exponential model selection. *IEEE Trans. Inform. Theory* **49** 2052-2060.

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

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In: *Handbook of Heavy Tailed Distributions in Finance* (S. T. Rachev, ed.) 329-384, Elsevier, Amsterdam.

Embrechts, P, McNeil, A. J. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk management: value at risk and beyond* (M. A. H. Dempster, ed.) 176-223, Cambridge Univ. Press, Cambridge.

Eubank, R. L., LaRiccia, V. N.and Rosenstein, R. B. (1987). Test statistics derived as components of Pearson's phi-squared distance measure. *J. Amer. Statist. Assoc.* **82** 816–825.

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

Genest, C. and A.-C. Favre (2006). Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, 11

Janic-Wróbleska, A., Kallenberg, W. C. M. and Ledwina, T. Detecting positive quadrant dependence and positive function dependence. *Insurance Math. Econom.* **34** 467-487.

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

Kallenberg, W. C. M. and Ledwina, T. (1999). Data-driven rank tests for independence. *J. Amer. Statist. Assoc.* **94** 285-301.

Lancaster, H.O. (1969). *The Chi-squared Distribution.* Wiley, New York.

McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, Princeton.

Nelsen, R. B. (1999). *An Introduction to Copulas.* Lecture Notes in Statistics **139** Springer-Verlag, New York.

Panchenko V. (2005). Goodness-of-fit test for copulas. *Phys. A* **355** 176-182.

Sklar, A., (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229-231.

Sklar, A. (1996). Random variables, distribution functions, and copulas – a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics* (L. Rüschendorf, B. Schweizer and M. D. Taylor, eds).1-14, Lecture notes monograph series **28**, Institute of Mathematical Statistics, Hayward, CA.

Yang, Y. and Barron, A. R. An asymptotic property of model selection criteria. *IEEE Trans. Inform. Theory* **44** 95-116.