Approximating Multivariate Distributions with Vines

Tim Bedford and Alireza Daneshkhah
Department of Management Science
University of Strathclyde
Glasgow, G1 1QE, UK

July 1, 2010

Abstract

In a series of papers, Bedford and Cooke used vine (or pair-copulae) as a graphical tool for representing complex high dimensional distributions in terms of bivariate and conditional bivariate distributions or copulae. In this paper, we show that how vines can be used to approximate any given multivariate distribution to any required degree of approximation. This paper is more about the approximation rather than optimal estimation methods. To maintain uniform approximation in the class of copulae used to build the corresponding vine we use minimum information approaches. We generalised the results found by Bedford and Cooke that if a minimal information copula satisfies each of the (local) constraints (on moments, rank correlation, etc.), then the resulting joint distribution will be also minimally informative given those constraints, to all regular vines. We then apply our results to modelling a dataset of Norwegian financial data that was previously analysed in Aas et al. (2009).

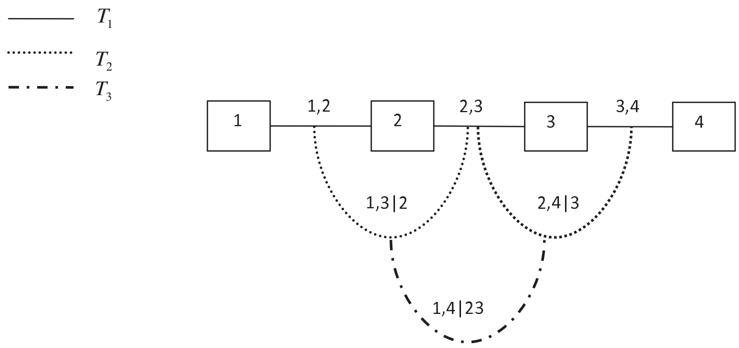


Figure 1: A regular vine with 4 elements

copula and copula density be $C_{jk|De}$ and $c_{jk|De}$ respectively. Let the marginal distributions F_i with densities $f_i, i = 1, \dots, n$ be given. Then the vine-dependent distribution is uniquely determined and has a density given by

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i) \prod_{j=1}^{n-1} \prod_{e(j,k) \in E_i} c_{jk|D_e}(F_{j|D_e}, F_{k|D_e}) \tag{2}$$

The existence of regular vine distributions is discussed in detail by Bedford and Cooke (2002).

The density decomposition associated with 4 random variables $\mathbf{X} = (X_1, \dots, X_4)$ with a joint density function $f(x_1, \dots, x_4)$ satisfying a copula-vine structure (this structure is called D-vine, see Kurowicka and Cooke, 2006, pp. 93) shown in Figure 1 with the marginal densities $f_1, \dots, f_4$ is

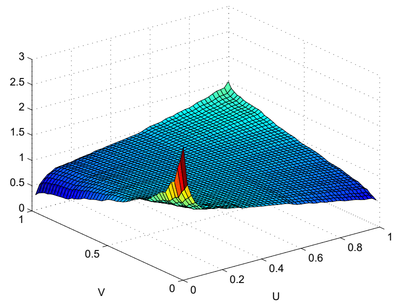
$$\begin{aligned} f_{1234}(x_1, \dots, x_4) \quad = \quad & \prod_{i=1}^{4} f(x_i) \times c_{12}\{F(x_1), F(x_2)\} c_{23}\{F(x_2), F(x_3)\} c_{34}\{F(x_3), F(x_4)\} \\ & \times c_{13|2}\{F(x_1 \mid x_2), F(x_3 \mid x_2)\} c_{24|3}\{F(x_2 \mid x_3), F(x_4 \mid x_3)\} \\ & \times c_{14|23}\{F(x_1 \mid x_2, x_3), F(x_4 \mid x_2, x_3)\} \end{aligned} \tag{3}$$

This formula can be derived for this case using the general expression

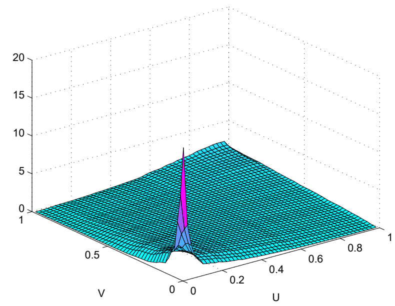
$$f(x, y) = f_X(x) f_Y(y) c(F_X(x), F_Y(y)),$$

or equivalently

$$f(x|y) = f_X(x) c(F_X(x), F_Y(y)),$$



3
2.5
2
1.5
1
0.5
0
1
0.5
0
V
0
0.2
0.4
0.6
0.8
1
U



20
15
10
5
0
1
0.5
0
V
0
0.2
0.4
0.6
0.8
1
U

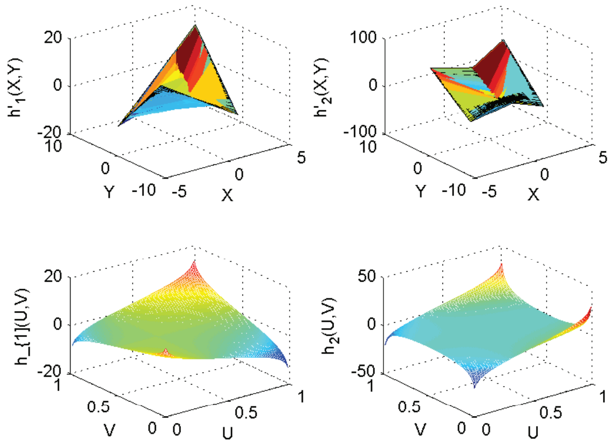


Figure 4: Plots of base functions and the corresponding functions on the copula domain

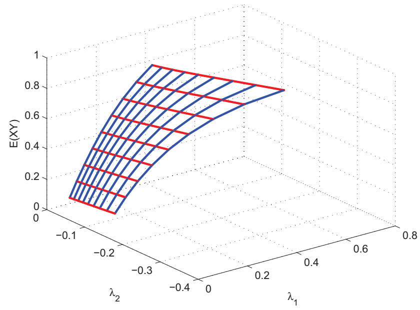


Figure 5: The presentation of $E(XY)$ as a function of λ_1 and λ_2

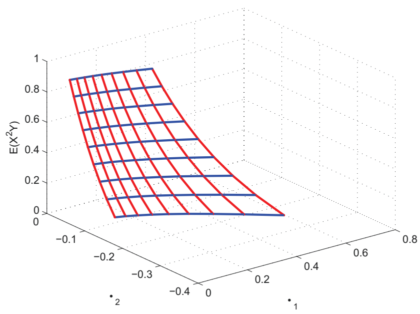
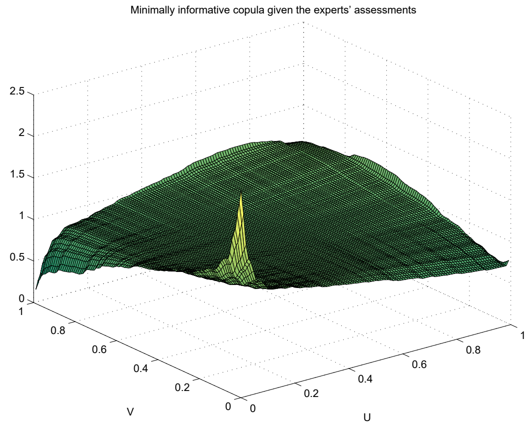


Figure 6: The presentation of $E(X^2Y)$ as a function of λ_1 and λ_2


Minimally informative copula given the experts' assessments
2.5
2
1.5
1
0.5
0
1
0.8
0.6
0.4
0.2
0
V
0
0.2
0.4
0.6
0.8
1
U

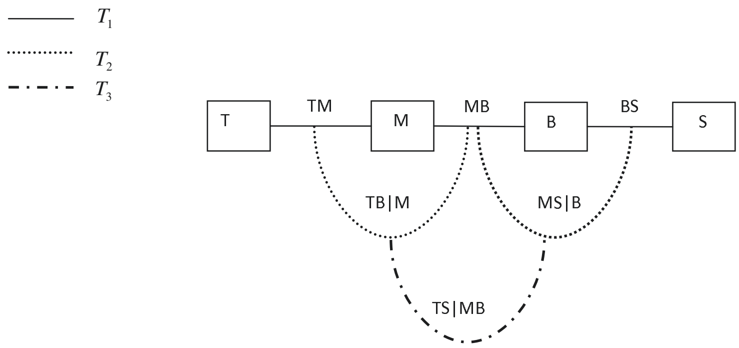


Figure 8: Selected vine structure for the Norwegian stock data set with 4 variables: Norwegian stock index (T), MSCI world stock index (M), Norwegian bond index (B) and SSBWG hedged bond index (S).

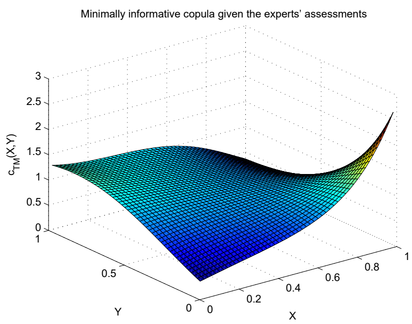
We illustrate the procedure by applying it to a financial data set.

Example 2 In this example, we use the same data set studied in Aas et al (2009). These are four time series of daily data: the Norwegian stock index (TOTX), the MSCI world stock index, the Norwegian bond index (BRIX) and the SSBWG hedged bond index, for the period from 0.4.01.1999 to 0.8.07.2003. We denote these four variables by T, M, B and S, respectively.

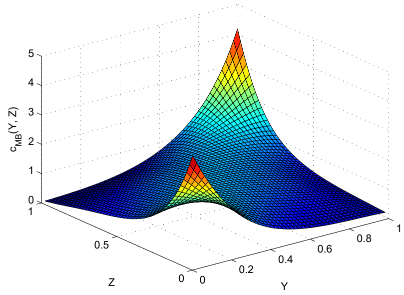
We want to generate vine approximation fitted to this data set to any given multivariate density using minimum information distribution. We select a similar vine structure with 4 elements shown in Figure 1 for this data presented in Figure 8. It should be noticed that, we can find the corresponding functions of the copula variables X, Y, Z and W associated with T, M, B, S, respectively, defined by $h_i(X,Y) = h_i'(F_1^{-1}(X), F_2^{-1}(Y))$, etc., and clearly these should also have the same specified expectation, that is, $E(h_i'(T,M)) = E(h_i(X,Y))$, etc. The minimum information copulae calculated in this example are derived based on the copula variables, X, Y, Z, W.

We first can construct a minimally informative copula between any two variables joining together in the first tree, T_1. As an example, we show the construction of a minimally informative copula between two variables M and T denoted by C_{TM} under the following constraints: $h_1'(M,T) = MT$, $h_2'(M,T) = TM^2$, $h_3'(M,T) = T^2M$ and $h_4'(M,T) = MT^3$. In other words, we use the Fourier copula of order 4 or a base with 4 elements to approximate this copula. We fix the values of the expectations of these functions as follows

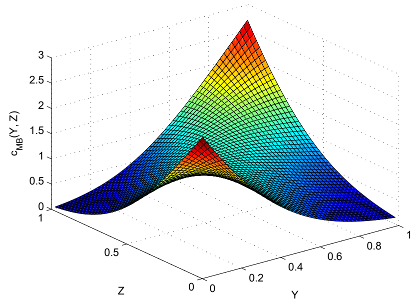
$$\alpha_1 = \frac{1}{1094}\sum_{i=1}^{1094} T_i M_i = 0.2314, \alpha_2 = \frac{1}{1094}\sum_{i=1}^{1094} T_i M_i^2 = 0.1497$$


Minimally informative copula given the experts' assessments
$c_{TM}(X,Y)$
3
2.5
2
1.5
1
0.5
0
1
0.5
0
Y
0
0.2
0.4
0.6
0.8
1
X

Minimally informative copula given the experts' assessments



Minimally informative copula given the experts' assessments



Bin	The constraints $(E[h'_1 \mid M], E[h'_2 \mid M], E[h'_3 \mid M])$	Lagrange multipliers $(\lambda_1, \lambda_2, \lambda_3)$
$0 < M < 0.1$	(0.1594, 0.0678, 0.1224)	(10.9383, -8.4123, -6.9916)
$0.1 < M < 0.2$	(0.1785, 0.0857, 0.1252)	(-4.2491, 8.1402, -5.2132)
$0.2 < M < 0.3$	(0.207, 0.1181, 0.1357)	(2.1269,-1.7432,-1.4931)
$0.3 < M < 0.4$	(0.1891,0.1032,0.1171)	(-7.2137,2.0704,1.9255)
$0.4 < M < 0.5$	(0.2587,0.1748,0.1653)	(-8.7337,5.2627,3.8922)
$0.5 < M < 0.6$	(0.2377,0.1538,0.1526)	(-12.5348,-1.6083,14.9014)
$0.6 < M < 0.7$	(0.2712,0.1802,0.1673)	(3.9591,-7.3273,3.5925)
$0.7 < M < 0.8$	(0.2595,0.1736,0.1618)	(7.3803,-15.482,8.9438)
$0.8 < M < 0.9$	(0.3156,0.2386,0.1945)	(9.2597,-9.2321,-0.4385)
$0.9 < M < 1$	(0.2626,0.2087,0.1618)	(-0.7429,-1.1895,0.6117)

Table 2: The constraints and corresponding Lagrange multipliers associated with the conditional minimal informative copula between $T \mid M \in (0,1)$ and $B \mid M \in (0,1)$ for each bin

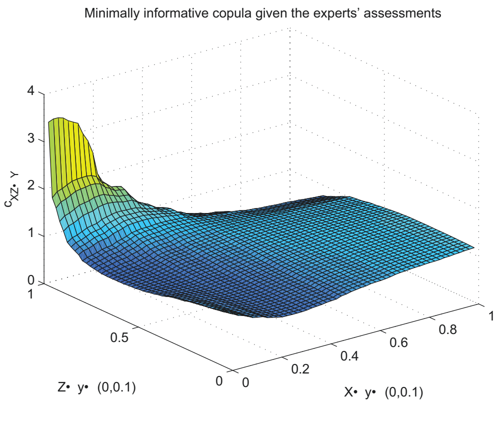


Figure 12: The minimally informative copula between $T \mid M \in (0, 0.1)$ and $B \mid M \in (0, 0.1)$ variables of Norwegian Stock data given $E[h'_1(T,B) \mid 0 < M < 0.1], E[h'_2(T,B) \mid 0 < M < 0.1], E[h'_3(T,B) \mid 0 < M < 0.1]$ constraints .

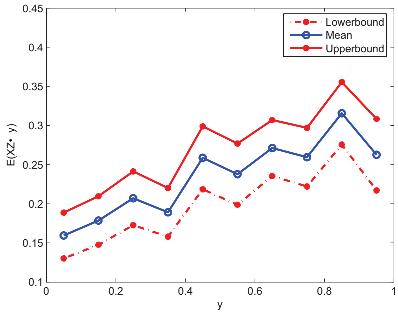


Figure 13: The changes of $E[h'_1(T,B) \mid 0 < M < 1]$ over the bins.

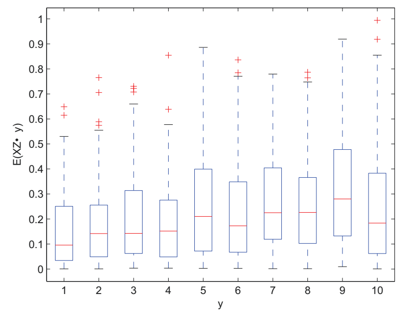


Figure 14: Box-plot demonstration of the $E[h'_1(T,B) \mid 0 < M < 1]$.

Table 2 shows the constraints and the corresponding Lagrange multipliers required to build conditional minimum information copula between $T \mid M \in (0,1)$ and $B \mid M \in (0,1)$ for 10 bins.

It is important to study the changes of the conditional expectation, $E[h'_1(T,B) \mid M]$ ($E[XZ \mid y]$) for different values of M or over the bins. Figure 13 shows this conditional expectation, $E[h'_1(T,B) \mid M]$, calculated from the minimum information copula $C(T \mid M, B \mid M)$ where M varies on $(0,1)$ along with the 95% confidence interval around the mean. As we can observe form this figure the changes of this measure is not...

The Box-plot demonstration of this conditional expectation, $E[h'_1(T,B) \mid 0 < M < 1]$ is illustrated in Figure 14. Similarly, we construct the conditional minimum information copula between $M \mid B$ and $S \mid B$ given the following constraints represented as the conditional expectations of some objective functions:

$$h'_1(M,S) = MS, \quad h'_2(M,S) = MS^2, \quad h'_3(M,S) = M^2S$$

Table 3 shows the constraints and the corresponding Lagrange multipliers required to build conditional minimum information copula between $M \mid B \in (0,1)$ and $S \mid B \in (0,1)$ for 10

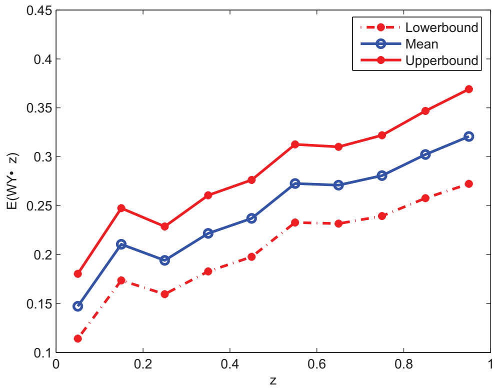


Figure 15: The conditional expectation $E[h_1'(M,S) \mid 0 < B < 1]$ derived from the minimally informative copula between $M \mid B \in (0,1)$ and $S \mid B \in (0,1)$ obtained above.

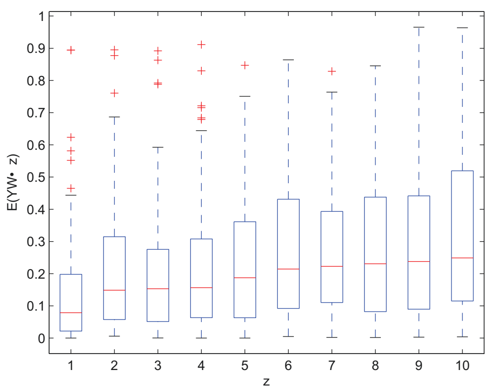


Figure 16: Box-plot demonstration of $E[h_1(M,S) \mid 0 < B < 1]$.

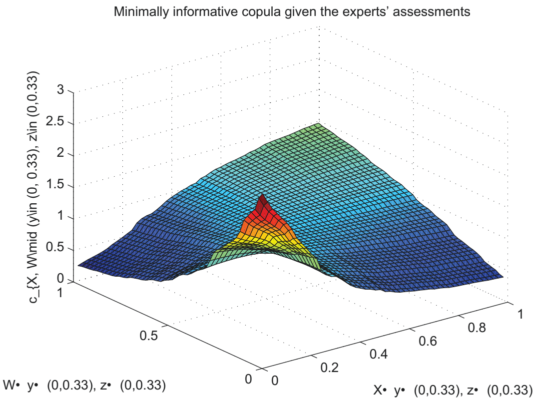


Figure 17: The minimally informative copula between $T \mid \{M \in (0.33), B \in (0, 0.33)\}$ and $S \mid \{M \in (0.33), B \in (0, 0.33)\}$ variables of Norwegian Stock data given $e_1 = E[h'_1(T, S) \mid \{M \in (0.33), B \in (0, 0.33)\}] = 0.394, e_2 = E[h'_2(T, S) \mid \{M \in (0.33), B \in (0, 0.33)\}] = 0.295, e_3 = E[h'_3(T, S) \mid \{M \in (0.33), B \in (0, 0.33)\}] = 0.3115$ constraints .

Bins	$(E[h'_1(T,S) \mid M,B]$ $E[h'_1(T,S) \mid M,B],$ $E[h'_1(T,S) \mid M,B])$	Lagrange multipliers $(\lambda_1, \lambda_2, \lambda_3)$
$0 < M < 0.33, 0 < B < 0.33$	(0.394,0.295,0.3115)	(5.0563,0.0806,-1.1935)
$0 < M < 0.33, 0.33 < B < 0.66$	(0.2995,0.1975,0.2192)	(5.8976,2.1862,-3.8)
$0 < M < 0.33, 0.66 < B < 1$	(0.2089,0.1346,0.1381)	(-0.5927,5.6017,0.1003)
$0.33 < M < 0.66, 0 < B < 0.33$	(0.2548,0.1731,0.1541)	(6.8429,2.8645,-6.5986
$0.33 < M < 0.66, 0.33 < B < 0.66$	(0.2459,0.1661,0.1623)	(10.2143,-3.9284,-1.6448)
$0.33 < M < 0.66, 0.66 < B < 1$	(0.2414,0.1643,0.1657)	(-5.0439,5.6452,3.3403)
$0.66 < M < 1, 0 < B < 0.33$	(0.2992,0.222,0.1976)	(0.3324,2.6365,1.6332)
$0.66 < M < 1, 0.33 < B < 0.66$	(0.2766,0.1942,0.1895)	(-0.0135,3.8203,-1.2379)
$0.66 < M < 1, 0.66 < B < 1$	(0.2163,0.1473,0.1334)	(16.729,-0.3679,-12.5079)

Table 4: The constraints and corresponding Lagrange multipliers associated with the conditional minimally informative copula between $T \mid (M, B)$ and $S \mid (M, B)$ for each bin.

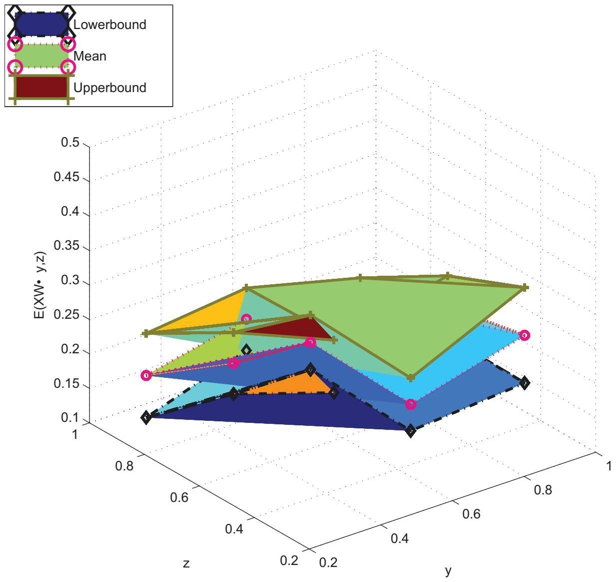


Figure 18: The conditional expectation $E[h_1(T,S) \mid 0 < M < 1, 0 < B < 1]$ derived from the minimally informative copula between $T \mid \{M \in (0,1), B \in (0,1)\}$ and $S \mid \{M \in (0,1), B \in (0,1)\}$ obtained above.

$1, 0 < B < 1]$ (the middle plane and recognised by "O" in the figure) and 95% confidence bound (we use "+" to display the upperbound, and "$\diamondsuit$" denotes the lowerbound) over the bins specified in Table 4.

6 Conclusion

In this paper, we present a novel method to approximate a multivariate distribution by any vine structure to any degree of approximation. Our approach uses the minimum information copulas that can be specified to any required degree of precision based on the data available. We prove rigourously that good approximation 'locally' guarantees good approximation globally. This approximation allows the use of a fixed finite dimensional family of copulas to be used in a vine construction, with the promise of a uniform level of approximation.In other words, we can use the same bases to approximate each copula in each tree of the corresponding vine.

However, a vine structure imposes no restrictions on the underlying joint probability distribution it represents, but this is crucial to investigate which vine structure is most appropriate. The choice of vine structure becomes more significant when we truncate class of copulae to make search strategy simpler. Therefore, the approximation of a multivariate