Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

THE VINE COPULA METHOD FOR REPRESENTING HIGH DIMENSIONAL DEPENDENT DISTRIBUTIONS: APPLICATION TO CONTINUOUS BELIEF NETS

Dorota Kurowicka
Roger M. Cooke

Dep. of Information, Technologies & Systems
Delft University of Technology
Mekelweg 4, 2628CD Delft
THE NETHERLANDS

ABSTRACT

High dimensional probabilistic models are often formulated as belief nets (BN's), that is, as directed acyclic graphs with nodes representing random variables and arcs representing "influence". BN's are conditionalized on incoming information to support probabilistic inference in expert system applications. For continuous random variables, an adequate theory of BN's exists only for the joint normal distribution. In general, an arbitrary correlation matrix is not compatible with arbitrary marginals, and conditionalization is quite intractable. Transforming to normals is unable to reproduce exactly a specified rank correlation matrix. We show that a continuous belief net can be represented as a regular vine, where an arc from node i to j is associated with a (conditional) rank correlation between i and j. Using the elliptical copula and the partial correlation transformation properties, it is very easy to conditionalize the distribution on the value of any node, and hence update the BN.

1 INTRODUCTION

Belief Nets (BN's) are Directed Acyclic Graphs (DAG's) representing influence between random variables. We would like to associate nodes in a BN with continuous univariate random variables and to interpret "influence" in terms of correlation. This should ideally be done in a way that does not impose intractable constraints, and which supports conditionalization. Simply associating arcs with correlations is unsuitable for two reasons; (i) the compatibility of marginal distributions and a specified product moment correlation is not easily determined, and (ii) the correlation matrix must be positive definite. To illustrate this latter constraint, consider the simple BN in Figure 4b. Variables 1 and 2 have correlation zero. If 1 is highly correlated with 3, then it is impossible that 2 is also highly correlated with 3, since this would require positive correlation between 1 and 2.

One option is to represent influence as rank correlation. Any rank correlation in the interval [-1, 1] is compatible with arbitrary continuous invertible marginals. We could transform the variables to standard normal, induce a product moment correlation structure using well known methods, and transform back to the original variables. The rank correlation thus induced would not be exactly equal the specified rank correlation, but would be in the right ball park. Exact replication of a given rank correlation matrix could be obtained with this method if the joint normal could realize every rank correlation matrix. This is not the case; indeed the rank correlation matrices of joint normal distributions are very sparse in the set of correlation matrices (Kurowicka and Cooke 2001). Of course, the problem of positive definiteness noted above would still be present.

In this paper we briefly describe the regular-vine-elliptical- copulae method for specifying dependences in high dimensional distributions (section 2). We then apply this method to the problem continuous BN's (sections 3 to 6). In section 7 we show how regular vines may be used to infer a BN from multivariate data.

2 VINES

2.1 Basic facts and concepts

A *vine* on N variables is a nested set of trees, where the edges of tree j are the nodes of tree j+1; j = 1,… ,N-2, and each tree has the maximum number of edges (Cooke 1997). A *regular* vine on N variables is a vine in which two edges in tree j are joined by an edge in tree j+1 only if these edges share a common node, j = 1,… ,N-2. There are (N-1)+(N-2)+ … +1 =N(N-1)/2 edges in a regular vine on N variables. Figure 1 shows a regular vine on 5 variables. The four nested trees are distinguished by the line style of the edges; tree 1 has solid lines, tree 2 has dashed lines, etc. The conditioned (before |) and conditioning (after |) sets associated with each edge are determined as follows: the variables reachable from a given edge are called the constraint set of that edge. When two edges are joined by an edge of the next tree, the intersection of the respective constraint sets are the conditioning variables, and the symmetric difference of the constraint sets are the conditioned variables. The regularity condition ensures that the symmetric difference of the constraint sets always contains two variables. Note that each pair of variables occurs once as conditioned variables.

We recall two generic vines, the *D-vine D(1,2,…n)* and C- -vine C(*1,2,…,n*), shown on Figures 1 and 2.

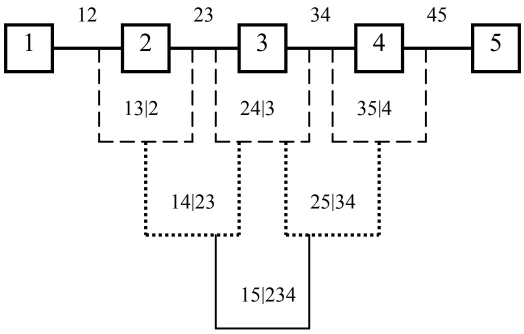

Figure 1: The D-Vine On 5 Variables D(12345) Showing Conditioned And Conditioning Sets

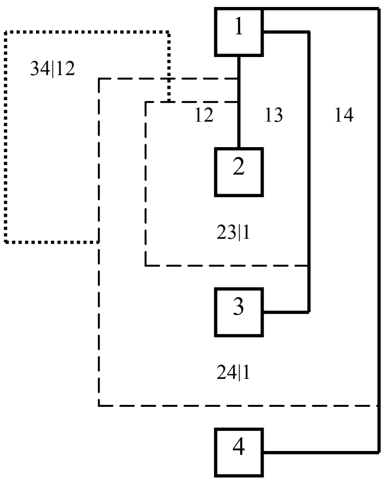

Figure 2: The C-Vine On 4 Variables C(1,2,3,4) Showing Conditioned And Conditioning Sets

Each edge in a regular vine may be associated with a constant conditional rank correlation (for j=1 the conditions are vacuous) and, using minimum information copulae, a unique joint distribution satisfying the vine-copulae specification with minimum information can be constructed and sampled on the fly (Cooke 1997). Moreover, the (constant conditional) rank correlations may be chosen arbitrarily in the interval [-1,1].

The edges of a regular vine may also be associated with partial correlations, with values chosen arbitrarily in the interval (-1,1). Using the well known recursive formulae (1) it can be shown that each such partial correlation regular vine uniquely determines the correlation matrix, and every full rank correlation matrix can be obtained in this way (Bedford and Cooke 2002). In other words, a regular vine provides a bijective mapping from $(-1,1)^{N(N-1)/2}$ into the set of positive definite matrices with 1's on the diagonal. One verifies that ρ_{ij} can be computed from the subvine generated by the constraint set of the edge whose conditioned set is {i, j} using the following recursive formulae

$$\rho_{12;3..n} = \frac{\rho_{12;3..n-1} - \rho_{1n;3...n-1}\rho_{2n;3...n-1}}{\sqrt{(1-\rho^2_{1n;3...n-1})(1-\rho^2_{2n;3...n-1})}} \ . \qquad (1)$$

When (X,Y) and (X,Z) are joined by the elliptical copula (see Figure 3) (Kurowicka et. al. 2000) and the conditional copula (Y,Z|X) does not depend on X, then the conditional correlation (Y,Z) given X does not depend on X and conditional product moment correlation of Y,Z given X is equal to partial correlation (Kurowicka and Cooke 2001).

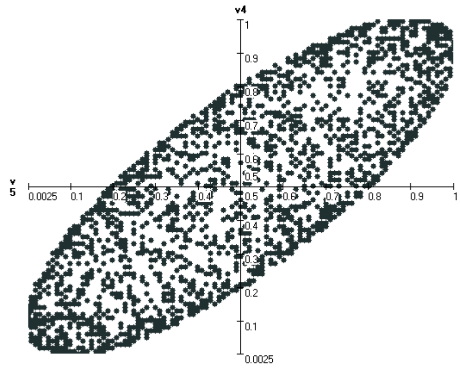

Figure 3: The Scatter Plot Of The Elliptical Copula With Correlation 0.8

Moreover, there exists very compact functional form of the conditional distribution using the elliptical copula.

Using elliptical copula vine structure can be uniquely associated with a full rank correlation matrix and can be converted into an on-the-fly sampling routine.

2.2 Updating vines

For a regular vine-rank correlation specification with elliptical copula updating with information is very simple. Since for elliptical copulae partial and conditional product moment correlations are equal then using the recursive formula (1) we can easily convert any partial correlation vine into any other partial correlation vine. We convert given vine to the C-vine with variable which we observe, say 1, as a root. Conditional correlations don't depend on a value of 1 then we drop the "1"s from all conditions and as a result we obtain C-vine with variable 2 as a root. We can convert this "updated vine" to any other regular vine, recalculate conditional rank correlations and sample this "updated" vine if desired.

3 BELIEF NETS

3.1 Basic concept

A finite valued *Belief Net* is a directed acyclic, graph, together with an associated set of probability tables. The graph consists of nodes and arcs. The nodes represent variables, which can be discrete or continuous. The arcs represent causal/influential or functional relationships between variables.

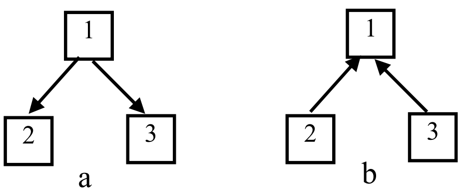

Figure 4: A Simple Example Of BN's

The graph in Figure 4a) tells us that variables 2 and 3 are conditional independent given variable 1. The message of the graph on Figure 4b) is that 2 and 3 are independent and a distribution of 1 given 2 and 3 is arbitrary.

If variables 1,2 and 3 on Figure 4a) take values "True" or "False" then two 2x2 probability tables must be specified (example shown in Table 1).

Table 1: An Example of Probability Tables Specification in BNs

2 \ 1	True	False
True	0.1	0.5
False	0.9	0.5

3 \ 1	True	False
True	0.6	0.7
False	04	0.3

Even for such a simple example, figuring out the right probabilities in the probability tables requires some work (e.g. statistical data or expert's opinions). For a large net with many dependences and nodes that can take more values this is almost impossible. For continuous random variables, an adequate theory of BN's exists only for the joint normal distribution. Despite of all these problems BN's are a very powerful tool that enables us to model uncertain events and describe dependence structure in an intuitive, graphical way.

3.2 Updating BNs

The main use of BNs is in situations that require statistical inference. If we know events that have actually been observed, we might want to infer the probabilities of other events, which have not yet been observed. Using Bayes Theorem it is then possible to update the values of all the other probabilities in the BN. Updating BN's is very complex involving arc reversing and addition but with the algorithm proposed by (Lauritzen and Spiegelhalter 1988) it is possible to perform fast updating in large BNs.

In the next section we show that a continuous belief net can be represented as a regular vine, where an arc from node i to j is associated with a (conditional) rank correlation between i and j.

4 ASSOCIATING A D-VINE WITH A BN

We number the nodes in a belief net 1,…,n.

Step 1 ***Sampling order.***
We construct a *sampling order* for the nodes, that is, an ordering such that all ancestors of node i appear before i in the ordering. A sampling order begins with a source node and ends with a sink node. Of course the sampling order is not in general unique.

Step 2 ***Factorize joint***
We first factorize the joint in the standard way following the sampling order. If the sampling order is 1,2,…n, write:

$$P(1,\ldots n) = P(1)P(2|1)P(3|12)\ldots P(n|1,2\ldots n-1).$$

Next, we underscore those nodes in each condition which are not necessary in sampling the conditioned variable. This uses (some of) the conditional independence relations in the belief net. For example, for the BN in Figure 5

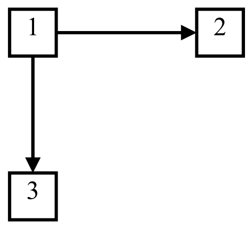

Figure 5: Simple BN

a sampling order is 1,2,3. Our factorization is

P(1)P(2|1)P(3|1$\underline{2}$).

For each term, we order the conditioning variables, i.e. the variables right of the "|", such that the underscored variables (if any) appear right-most and the non-underscored variables left-most.

Step 3 ***Quantify D-vine for node n***
Suppose the last term looks like:

P(n | n-1,n-3,…$\underline{\text{n-2, 3,2,1}}$).

Construct the D-vine with the nodes in the order in which they appear, starting with n (left) and ending with the last underscored node (if any).

If the D-vine D(n-1, n-3,…1) is given, the D-vine D(n, n-1,…1) can be obtained by adding the edges:

(n, n-1), (n, n-3| n-1)….(n, $\underline{1}$ | n-1,…$\underline{2}$).

For any underscored node $\underline{k}$, we have

(n ⊥ $\underline{k}$ | all non-underscored nodes ∪ any subset of underscored's not including. k).

The conditional correlation between n and an underscored node will be zero.

For any non-underscored node j, the bivariate distribution

(n,j | non-underscored nodes before j)

will have to be assessed. The conditioned variables (n,j) correspond to an arc in the belief net.

Write these conditional bivariates next to the corresponding arcs in the belief net. Note that we can write the (conditional) correlations associated with the incoming arcs for node n without actually drawing the D-vines. If the last factor is P(5|1,2,3,$\underline{4}$), we have incoming arcs (5,1), (5,2) and (5,3) which we associate with conditional correlations (5,1), (5,2|1) and (5,3|12).

Step 4 ***Quantify D- vine for node n-1, for node n-2 etc.***
Proceed as in step 3 for nodes 1,2,…n-1. Notice that the order of these nodes need not be the same as in the previous step. Continue until we reach the D-vine D(12) or until the order doesn't change in smaller subvines. i.e, if for node 4 the D-vine is D(4321) and for node 3 it is D(321) then we can stop with node 4; or better, we can quantify the vine D(321) as a subvine of D(4321).

Step 5 ***Construct partial correlation D-vine (1,…n)***

As a result of steps 1 - 4 each arc in the belief net is associated with a (conditional) bivariate distribution. These conditional distributions do not necessarily agree with the edges in D(1…n) since the orders in the different steps may be different. However, if the conditional bivariate distributions are given by partial correlations, then given D(1…k) we can compute D(π(1)…π(k)) where π ∈ k! Is a permutation of 1,…k. Repeatedly using this fact, we compute the partial correlations for D(1…n).

Since the values of conditional correlations on regular (sub)vines are algebraically independent and uniquely determine the correlation (sub)structure, the above algorithm leads to an assignment of correlations and conditional correlations to arcs in a belief net which are algebraically independent and which, together with the "zero edges" in the corresponding D-vines, uniquely determine the correlation matrix. Example 1 (Figures 6-10) illustrates the above steps.

Example 1

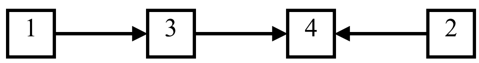

Figure 6: Example 1; BN

Sampling order: 1234

Factorization: P(1)P(2|$\underline{1}$)P(3|1$\underline{2}$)P(4|23$\underline{1}$)

D-vine 4231:

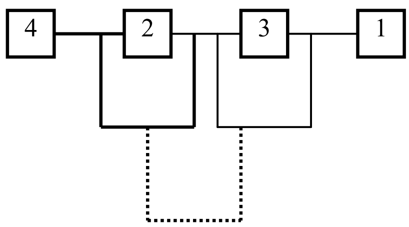

Figure 7: Example 1; Step 2,3

The dotted edge has partial correlation zero, the bold edges correspond to (4,2) and (4,3|2). These are written on the belief net and must be assessed.

We now consider the tem P(3|1$\underline{2}$). The order is different than for the term P(4 | 23$\underline{1}$). We construct D(312):

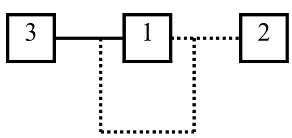

Figure 8: Example 1; Step 4

The dotted edges have partial correlation zero, the bold edge must be assessed, it is added to the belief net. The belief net is now quantified:

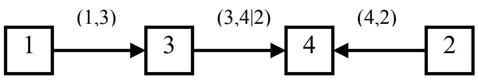

Figure 9: Example 1; BN With Conditional Correlations

With the partial correlations in D(312) we compute using the recursive relations D(231). In fact, we find $\rho_{23} = \rho_{23;1} = 0$. We now have D(4231) which corresponds to the belief net:

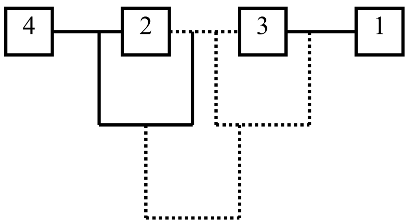

Figure 10: D-vine Corresponding to Example 1

Example 2
The steps for this example are illustrated in Figures 11-15.

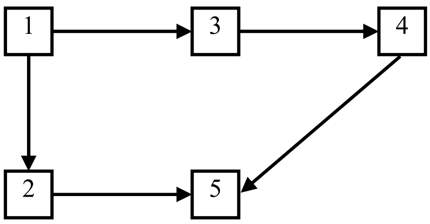

Figure 11: Example 2; BN

Sampling order: 12345

Factorization: P(1)P(2|1)P(3|12)P(4|321)P(5|2413)
D(52413):

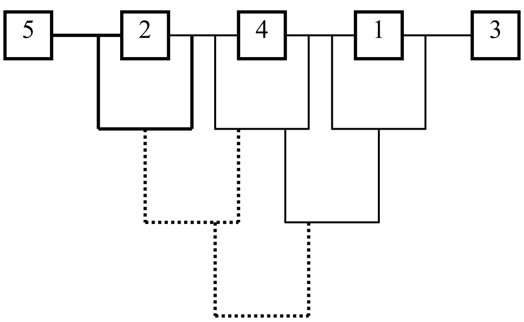

Figure 12: Example 2; Step 2,3

D(4321):

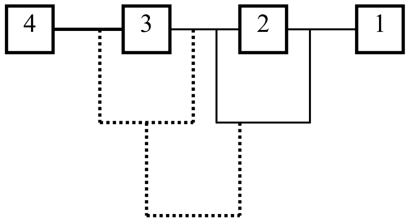

Figure 13: Example 2; Step 4

D(312):

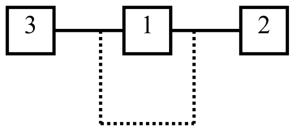

Figure 14: Example 2; Step 4 (continued)

We don't need to draw D(21). The Belief net becomes:

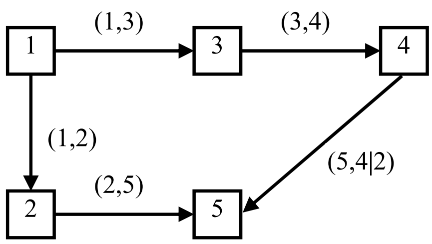

Figure 15: Example 2; BN With Conditional Correlations

Example 3
We may omit drawing all the vines by noting that we can identify the (conditonal) correlations which must be specified on the BN directly in steps 3 and 4 of the above algorithm. Thus consider the BN in Figure16.

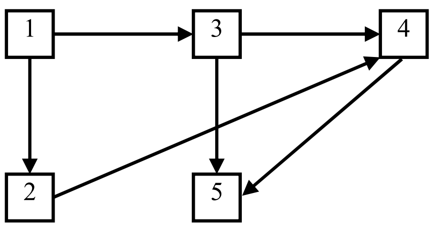

Figure 16: Example 3; BN

Sampling order: 12345

Factorization: P(1)P(2|1)P(3|12)P(4|231)P(5|3412).
This enables us to write the necessary conditonal correlations directly on the BN (see Figure 17):

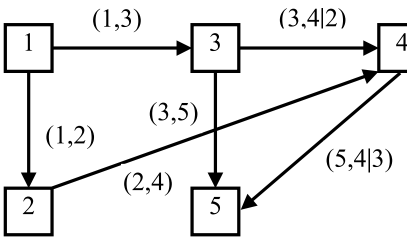

Figure 17: Example 3; BN With Conditional Correlations

5 CANONICAL VINE

We could also use the C-vine in the same way as the D-vine in the above procedure. The only problem is that it doesn't draw nicely. With the C-vine, the procedure is different according to whether we add a new variable to the end or the beginning of the vine. For example, suppose we have C(1234) and we want to add variable 5 *at the end* to construct C(12345). We have to add edges corresponding to the conditions: (51), (5,2|1), (5,3|12), (5,4|123). This is the same as for the D-vine D(51234). However if we want to add 5 to the beginning, to form C(51234) then we have to add conditions (51), (52), (53), and (54) and we have to add 5 to the set of conditioning variables in each of the nodes of C(1234).
As mentioned above, C-vines are most useful for actual sampling and for conditionalization. Hence in using belief nets, we will be doing lots of vine transformations. The elliptical copula will not create errors, whereas for some other copula, errors in approximating conditional as partial correlation will cumulate.

Example 4
In this example we use C-vine to quantify the belief net. The above steps are illustrated in Figures 18-20.

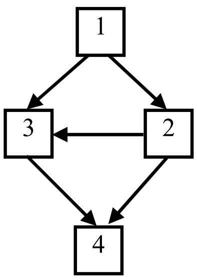

Figure 18: Example 4; BN

Sampling order: 12345

Factorization: P(1)P(2|1)P(3|2 1)P(4|2 3 1)

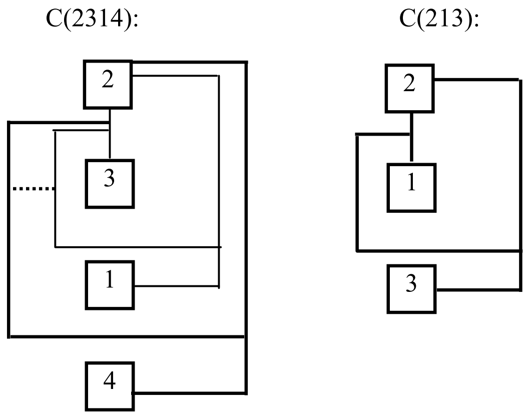

Figure 19: Example 4; Step 2,3,4 With C-vine

As a result we obtain:

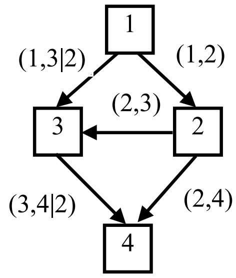

Figure 20: Example 4; BN With Conditional Correlations

6 FROM VINES TO BELIEF NETS

Obviously, from a given D-vine or C-vine we can't uniquely construct a belief net, as the order of the sub vines may change. However, in some simple cases we can invert the algorithm of Section 2. Dotted edges indicate zero correlation.

Example 5
Let us consider the D-vine in Figure 21.

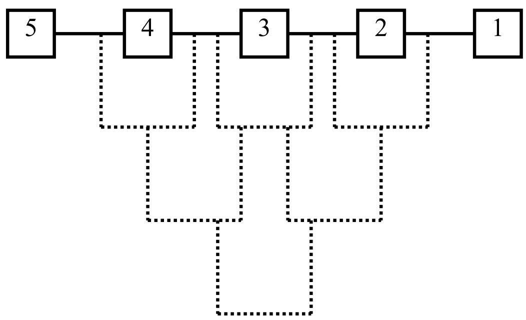

Figure 21: Example 5; D-vine

This corresponds to a factorization:

$$P(1)\,P(2\mid 1)\,P(3\mid 2\,\underline{1})\,P(4\mid 3\,\underline{2}\,\underline{1})\,P(5\mid 4\,\underline{3}\,\underline{2}\,\underline{1}).$$

The belief net is shown in Figure 22

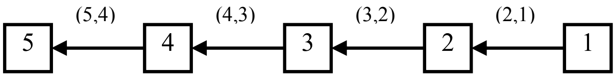

Figure 22: BN With Conditional Correlations Corresponding to Example 5

Example 6
Let us consider the D-vine in Figure 22.

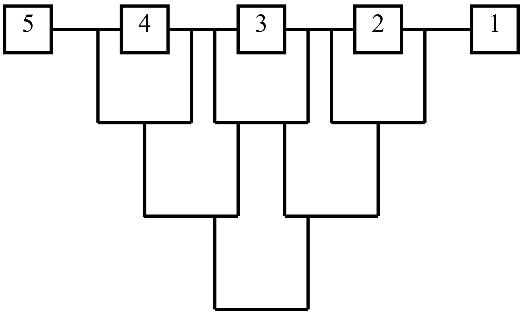

Figure 23: Example 6; D-vine

The factorization is

$$P(1)\,P(2\mid 1)\,P(3\mid 21)\,P(4\mid 321)\,P(5\mid 4321).$$

The belief net is shown in Figure 24:

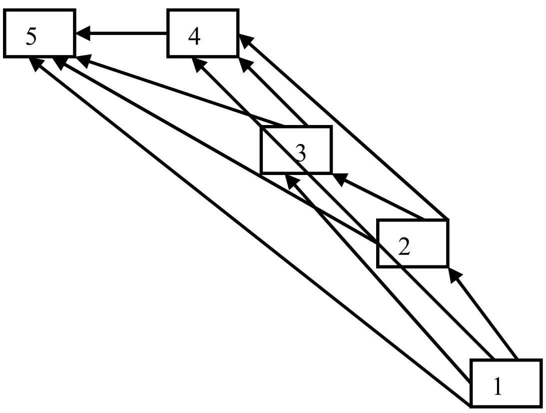

Figure 24: Belief Net With 5 Nodes Corresponding to the D-vine

Not very interesting. We don't add the conditions as they correspond to all the edges in the D-vine.

Example 7
Let us consider the D-vine in Figure 25.

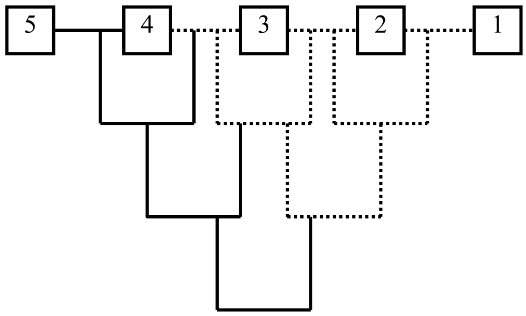

Figure 25: Example 7; D-vine

The factorization is

$$P(1)\,P(2\mid \underline{1})\,P(3\mid \underline{21})\,P(4\mid \underline{321})\,P(5\mid 4321).$$

The belief net is shown in Figure 26.

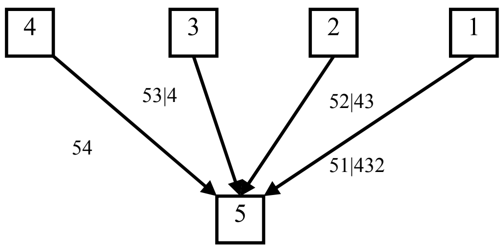

Figure 26: BN With Conditional Correlations Corresponding To Example 7

Example 8
Let us consider the D-vine in Figure 27.

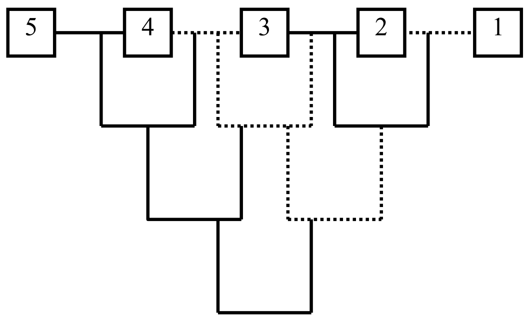

Figure 27: Example 8; BN

Factorization:

$$P(1)\,P(2\mid \underline{1})\,P(3\mid 21)\,P(4\mid \underline{321})\,P(5\mid 4321).$$

The belief net is shown in Figure 28.

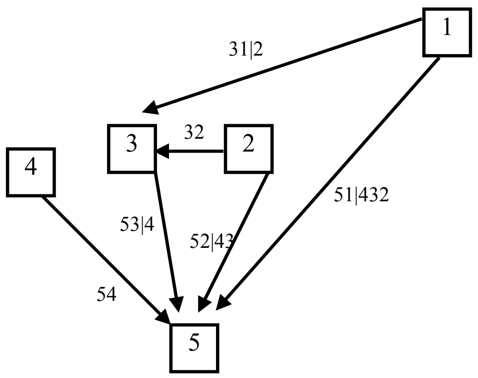

Figure 28: BN With Conditional Correlations Corresponding To Example 8

And so on.

7 FROM CORRELATION MATRIX TO BN

The interesting problem arises when we want to reconstruct the graphical dependence structure from the data. The correlation matrix can be calculated and cells significantly different then zero selected. The correlation matrix calculated from the date is changed to one with zeros in places where correlations don't differ significantly from zero. (Note that this procedure can be quite difficult since the changed matrix must satisfy positive definiteness constraint). Then we can calculate correlations on the C or D-vine and determine the belief net corresponding to it. The easiest way would be to take belief net as on Figure 24, calculate all partial correlations on the D-vine and assign them to arcs of the belief net. This is simple but not too smart since we would like to recover from the data the most conditional indifference relationships and reduce maximally number of arcs in the belief net. This can be done by finding the right ordering of variables that gives us maximal number of partial correlations on the C- vine equal to zero.
The following procedure can be used:

- Choose a variable with maximum number of zeros in the correlation matrix (say A) as a first in the ordering. (It is obvious that this way we maximize number of zeros in the first tree of canonical vine on Figure 2).
- Calculate a matrix of partial correlations of the first order given this chosen first variable. (We denote this matrix $A_{;1} = [\rho_{ij;1}]$). Using formula (1) we immediately see that the matrix $A_{;1}$ has zero in every cell where ρ_{1j}=0 and ρ_{ij}=0, i=2,…,n. (Notice that $\rho_{ij;1}$=0 also when $\rho_{ij}=\rho_{1i}\rho_{1j}$)
- As a second in the ordering choose variable which has the most zeros in the matrix $A_{;1}$. This way the maximum number of zeros will be obtained in the second tree of the canonical vine etc…

Of course the above procedure doesn't give the unique ordering but always produces a canonical vine with maximum number of partial correlations equal to zero.

EXAMPLE 9
Let us consider the following matrix

1	0.5	0	0	0
0.5	1	0.1	0	0
0	0.1	1	0.7	0
0	0	0.7	1	0.2
0	0	0	0.2	1

Following the above procedure we find new ordering of variables [1 5 2 4 3] and the C-vine

ρ_{15}	ρ_{12}	ρ_{14}	ρ_{13}		0	0.5	0	0
	$\rho_{25;1}$	$\rho_{45;1}$	$\rho_{35;1}$	=		0	0.2	0
		$\rho_{24;15}$	$\rho_{23;15}$				0	0.1155
			$\rho_{34;152}$					0.7192

Now we build the belief net corresponding to this vine:

- Start with the last variable , that is 3. All partial correlations on the C-vine involving 3 as a conditioned variables appear in the last column of the above matrix.
- Add arrow to the node 3 from each variable which appear with 3 in conditioned set of non-zero partial correlation. (Correlations 23;15 and 34;152 are non-zero hence 3 gets arrows from nodes 2 and 4).
- Proceed the same way with the next in backward ordering variable, that is 4. Draw arrows to 4 from each variable which appear in conditioned set of non-zero correlation in third column of the above matrix. (Correlation 45;1 is non-zero hence 4 gets arrow from 5)
- In the end correlation 12 is nonzero hence 2 gets arrow from 1.

We obtain the following belief net shown in Figure 29.

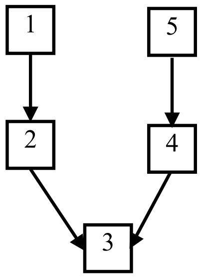

Figure 29: BN Corresponding To Example 9

CONCLUSIONS

1. Using the elliptical copula and regular vines, we have shown how an arbitrary belief net with arbitrary continuous univariate distributions at the nodes can be quantified. Influence is represented as (conditional) rank correlation. The (conditional) correlations are algebraically independent and, in combination with the (conditional) indifference statements implied by the belief net, uniquely determine the correlation structure.
2. Sampling and conditionalization are easily performed by transforming to the appropriate C-vine.
3. Finally, we show how the regular-vine-elliptical-copula approach can be used to infer a belief net from an empirical correlation matrix.

REFERENCES

T.J. Bedford and R.M. Cooke, 2002, Vines - a new graphical model for dependent random variables, Accepted to *Ann. of Statistics*.

R.M. Cooke,1997, Markov and entropy properties of tree and vines-dependent variables, *Proceedings of the ASA Section of Bayesian Statistical Science*.

D. Kurowicka, J. Misiewicz, and R.M.Cooke, 2000, Elliptical copulae, *Proc. of International Conference on Simulations- Monte Carlo*, 201-214.

D. Kurowicka, and R.M. Cooke, 2001, Conditional, Partial and Rank Correlation for Elliptical Copula; *Dependence Modeling in Uncertainty Analysis, Proc. of ESREL 2001*.

S.L. Lauritzen, 1996, *Graphical Models*, Clarendon Press, Oxford.

S.L. Lauritzen, D.J. Spiegelhalter, 1988, Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society*, Series B, vol. 50, pp.157-224.

J.Q. Smith, 1988, Statistical principals on graphs, in *Proceedings Influence Diagrams for Decision Analysis, Inference and Prediction,* Smith and Oliver eds, Wiley, Chichester.

AUTHOR BIOGRAPHIES

DOROTA KUROWICKA studied mathematics at Gdansk University in Gdansk and specialized in numerical methods. Her Masters thesis was completed in 1990. She performed Ph.D.research at the Gdynia Maritime academy and the Delft University of Technology, and received her PhD from Delft in 2000. Since then she is employed as Assistant Professor at the Delft University of Technology, Department of Mathematics. Her current research is carried out in the area of uncertainty and sensitivity analysis.

ROGER COOKE received his bachelors and doctoral degrees at Yale University in mathematics and philosophy. He taught logic and philosophy of science at the Yale University, the University of Amsterdam and the Delft University of Technology. In 1994 he was appointed to the chair of Applications of Decision Theory in the Department of Mathematics at Delft. His current interests are methodological aspects of risk analysis.