Vines in Overview
Invited Paper Third Brazilian Conference on Statistical Modelling in Insurance and Finance Maresias, March 25-30, 2007

R.M. Cooke, O. Morales, D. Kurowicka
Resources for the Future and
Department of Mathematics, Delft University of Technology
Cooke@rff.org

Abstract

Regular vines are a graphical tool for representing complex high dimensional distributions as bivariate and conditional bivariate distributions. This paper reviews previous results on vines, and suggests a new strategy for using regular vines in inferring model structure, where the relations between variables is not linear.
Keywords: correlation, graphs, positive definite matrix, matrix completion, model inference.
AMS classifications: 60H30, 6209, 62D05, 62H05, 15A48, 15A99

1 Introduction

Vines are graphical models for building up high dimensional distributions from conditional bivariate pieces. When univariate continuous margins are specified and when conditional copulae are assigned to the nodes of a regular vine, a routine can be specified for sampling the distribution. This provides the basis for a theory of continuous non-parametric Bayesian belief nets, and offers directions for model inference based on arbitrary functional relations between random variables.

This paper reports without proofs the main results about vines, sketches new research directions, and gives an illustrative example of the use of vines in capturing complex dependence structures. The example is used to illustrate the software tools which help understand complex distributions. For information on copulae, we refer to [16] and the references therein. The vine sampling algorithms are explored in [19]; the results regarding model learning in section 6 are new, and extend the ideas in [17].

2 Definitions

A partial correlation can be defined in terms of partial regression coefficients.

Consider variables X_i with zero mean and standard deviations σ_i, $i = 1, \ldots, n$, and let the numbers $b_{12;3,\ldots,n}, \ldots, b_{1n;2,\ldots,n-1}$ minimize

$$E\left((X_1 - b_{12;3,\ldots,n}X_2 - \ldots - b_{1n;2,\ldots,n-1}X_n)^2\right).$$

Definition 2.1 (Partial correlation)

$$\rho_{12;3,\ldots,n} \quad = \quad sgn(b_{12;3,\ldots,n})\,(b_{12;3,\ldots,n}b_{21;3,\ldots,n})^{\frac{1}{2}}\ ,\ \ etc.$$

Equivalently we could define the partial correlation as

$$\rho_{12;3,\ldots,n} \quad = \quad -\frac{C_{12}}{\sqrt{C_{11}C_{22}}},$$

where $C_{i,j}$ denotes the $(i,j)th$ cofactor of the correlation matrix. The partial correlation $\rho_{12;3,\ldots,n}$ can be interpreted as the correlation between the orthogonal projections of X_1 and X_2 on the plane orthogonal to the space spanned by $X_3,\ldots,X_n$.

Partial correlations can be computed from correlations with the following recursive formula [25].

$$\rho_{12;3,\ldots,n} \quad = \quad \frac{\rho_{12;3,\ldots,n-1} - \rho_{1n;3,\ldots,n-1}\cdot\rho_{2n;3,\ldots,n-1}}{\sqrt{1-\rho^2_{1n;3,\ldots,n-1}}\sqrt{1-\rho^2_{2n;3,\ldots,n-1}}}. \tag{1}$$

Definition 2.2 (Multiple correlation) *The* **multiple correlation** $R_{1\{2,\ldots,n\}}$ *of variables* X_1 *with respect to* $X_2,...,X_n$ *is*

$$1-R^2_{1\{2,\ldots,n\}} = \frac{D}{C_{11}}$$

where D is the determinant of the correlation matrix. It is the correlation between X_1 and the best linear predictor of X_1 based on $X_2,...,X_n$.

In [11] it shown that $R_{1\{2,\ldots,n\}}$ is non negative and satisfies:

$$1-R^2_{1\{2,\ldots,n\}} \quad = \quad (1-\rho^2_{1n})(1-\rho^2_{1,n-1;n})(1-\rho^2_{1,n-2;n-1,n})...(1-\rho^2_{12;3\ldots n}). \tag{2}$$

It follows from [11], that $R_{1\{2,\ldots,n\}}$ is invariant under permutation of $\{2,...,n\}$ and

$$D = \left(1-R^2_{1\{2,\ldots,n\}}\right)\left(1-R^2_{2\{3,\ldots,n\}}\right)...\left(1-R^2_{n-1\{n\}}\right); \tag{3}$$

Of course $R_{n-1\{n\}} = \rho_{n-1,n}$.

Definition 2.3 (Relative information, Mutual information) *Let f and g be densities on $\mathbb{R}^n$ with f absolutely continuous with respect to g;*

- *the* **relative information** *of f with respect to g is*

$$I(f|g) = \int_1 \ldots \int_n f(x_1\ldots x_n)\, ln(\frac{f(x_1\ldots x_n)}{g(x_1\ldots x_n)})dx_1\ldots dx_n$$

- *the* **mutual information** *of f is*

$$MI(f) = I(f \mid \Pi^n_{i=i} f_i)$$

where f_i is the $i-th$ univariate marginal density of f and $\Pi^n_{i=1}f(x_1\ldots x_n)$ is the independent distribution with univariate margins $\{f_i\}$.

Relative information is also called the Kullback Leibler information and the directed divergence. The mutual information is also called the information proper. The mutual information will be used to capture general dependence in a set of multivariate data. We do not possess something like an 'empirical mutual information', rather it must be estimated with kernel estimators, as suggested in [9]. For some copulae, the mutual information can be expressed in closed form [16]:

Theorem 2.1 *Let g be the elliptical copula with correlation ρ, then the mutual information of g is*

$$1 + ln2 + ln(\pi\sqrt{1-\rho^2}).$$

Let h be the diagonal band copula with vertical bandwidth parameter $1-\alpha$, then the mutual information of h is

$$-ln(2^{|\alpha|}(1-|\alpha|))$$

Note that the mutual information of the elliptical copula with zero correlation is not zero, reflecting the fact that zero correlation in this case does not entail independence.

3 Vines

Graphical models called *vines* were introduced in [3, 2, 14]. A vine $\mathcal{V}$ on n variables is a nested set of connected trees $\mathcal{V} = \{T_1, ..., T_{n-1}\}$ where the edges of tree j are the nodes of tree $j+1$, $j = 1, ..., n-2$. 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, $j = 1, ..., n-2$. The formal definitions follow.

Definition 3.1 (Regular vine) *$\mathcal{V}$ is a regular vine on n elements if*

1. $\mathcal{V} = \{T_1, \ldots, T_{n-1}\}$,

2. *T_1 is a connected tree with nodes $N_1 = \{1, \ldots, n\}$, and edges E_1;*
for $i = 2, \ldots, n-1$ T_i is a tree with nodes $N_i = E_{i-1}$. $E(\mathcal{V})$ denotes the set of edges of $\mathcal{V}$

3. (proximity) *for $i = 2, \ldots, n-1$, $\{a, b\} \in E_i, \#a\triangle b = 2$ where $\triangle$ denotes the symmetric difference.*

A regular vine is called a *canonical* or *C vine* if each tree T_i has a unique node of degree[1] $n-i$, hence has maximum degree. A regular vine is called a *D-vine* if all nodes in T_1 have degree not higher than 2 (see Figure 1). There are $n(n-1)/2$ edges in a regular vine on n variables. An edge in tree T_j is an unordered pair of nodes of T_j, or equivalently, an unordered pair of edges of T_{j-1}. By definition, the *order* of an edge in tree T_j is $j-1$, $j = 1, ..., n-1$.

The above definition is explained using the D-vine in Figure 1. There are 3 trees in the D-vine on 4 variables

$$\begin{aligned}
V &= \{T_1, T_2, T_3\}, \\
T_1 &= (N_1, E_1): \; N_1 = \{1, 2, 3, 4\}; \; E_1 = \{(1, 2), (2, 3), (3, 4)\}, \\
T_2 &= (N_2, E_2): \; N_2 = E_1; \; E_2 = \{((1, 2), (2, 3)), ((2, 3), (3, 4))\}, \\
T_3 &= (N_3, E_3): \; N_3 = E_2; \; E_3 = \{(((1, 2), (2, 3)), ((2, 3), (3, 4)))\}.
\end{aligned}$$

A regular vine is just a way of identifying a set of conditional bivariate constraints. The conditional bivariate constraints associated with each edge are determined as follows: the variables reachable from a given edge via the membership relation 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 differences of the constraint sets are the *conditioned variables*. More precisely the constraint, the conditioning and the conditioned set of an edge can be defined as follows:

Definition 3.2 *1. For $e \in E_i, i \leq n-1$, the* **constraint set** *associated with e is the* **complete union** *U_e^* of e, that is, the subset of $\{1, \ldots, n\}$ reachable from e by the membership relation.*

[1] The degree of node is the number of edges attached to it.

2. *For* $i = 1, \ldots, n-1, e \in E_i$, *if* $e = \{j, k\}$ *then the* **conditioning set** *associated with* e *is*

$$D_e = U_j^* \cap U_k^*$$

and the **conditioned set** *associated with* e *is*

$$\{C_{e,j}, C_{e,k}\} = \{U_j^* \setminus D_e,\ U_k^* \setminus D_e\}.$$

Note that for $e \in E_1$, the conditioning set is empty. One can see that the order of an edge is the cardinality of its conditioning set. For $e \in E_i, i \leq n-1, e = \{j, k\}$ we have $U_e^* = U_j^* \cup U_k^*$.

The following propositions are proved in [2, 14, 18]:

Proposition 3.1 *Let* $\mathcal{V} = \{T_1, ...T_{n-1}\}$ *be a regular vine, then*

1. *the number of edges is* $\frac{n(n-1)}{2}$,
2. *each conditioned set is a doubleton, each pair of variables occurs exactly once as a conditioned set,*
3. *if two edges have the same conditioning set, then they are the same edge.*

Remark 3.1 *When* $e \in E(\mathcal{V})$ *for a regular vine* $\mathcal{V}$, *we let* D_e *denote the conditioning set associated with* e *and* $\{e_1, e_2\}$ *the conditioned set.*

Definition 3.3 (m-child; m-descendent) *If node* e *is an element of node* f, *we say that* e *is an* **m-child** *of* f; *similarly, if* e *is reachable from* f *via the membership relation:* $e \in e_1 \in ... \in f$, *we say that* e *is an* **m-descendent** *of* f.

Proposition 3.2 *For any node* K *of order* $k > 0$ *in a regular vine*[2], *if variable* i *is a member of the conditioned set of* K, *then* i *is a member of the conditioned set of exactly one of the m-children of* K, *and the conditioning set of an m-child of* K *is a subset of the conditioning set of* K.

The edges of a regular vine may be associated with partial correlations, with values chosen arbitrarily in the interval $(-1, 1)$ in the following way:

To each $e \in E(\mathcal{V})$ we associate

$$\rho_{e_1,e_2;D_e}.$$

where $\rho_{e_1,e_2;D_e} = \rho_{e_1,e_2}$ if D_e is vacuous. The result is called a *partial correlation vine.*

Theorem 3.1 [2] shows that each such partial correlation vine specification uniquely determines the correlation matrix, and every full rank correlation matrix can be obtained in this way. In other words, a regular vine provides a bijective mapping from $(-1, 1)^{\binom{n}{2}}$ into the set of positive definite matrices with 1's on the diagonal.

Theorem 3.1 *For any regular vine on* n *elements there is a one to one correspondence between the set of* $n \times n$ *positive definite correlation matrices and the set of partial correlation specifications for the vine.*

[2] Equivalently one can formulate this lemma for edges of $\mathcal{V}$.

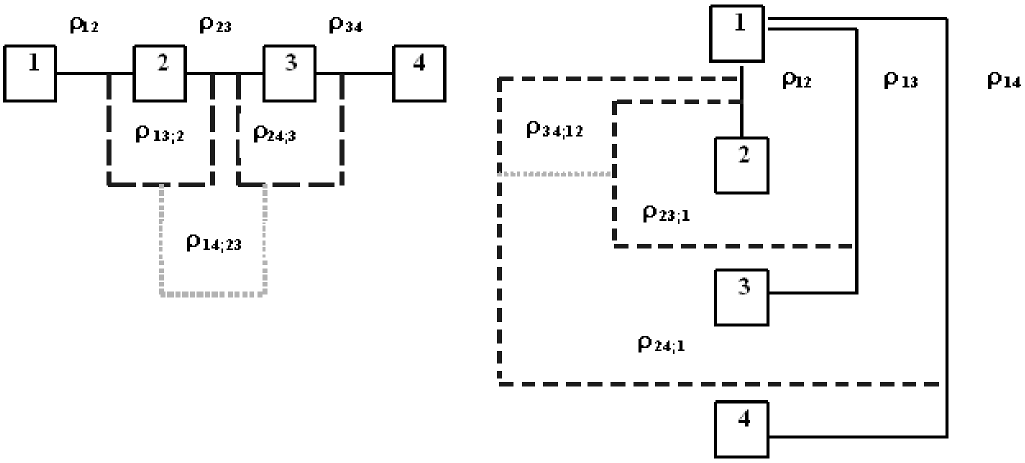

Figure 1: Partial correlations D-vine (left) and canonical vine (right) on 4 variables.

All assignments of the numbers between -1 and 1 to the edges of a partial correlation regular vine are consistent, and all correlation matrices can be obtained this way.

One verifies that the correlation between ith and jth variables can be computed from the sub-vine generated by the constraint set of the edge whose conditioned set is $\{i, j\}$ using recursive the formulae (1), and the following lemma.

Lemma 3.1 *If $z, x, y \in (-1, 1)$, then also $w \in (-1, 1)$, where*

$$w = z\sqrt{(1-x^2)(1-y^2)} + xy.$$

A regular vine may thus be seen as a way of picking out partial correlations which uniquely determine the correlation matrix and which are algebraically independent. The partial correlations in a partial correlation vine need not satisfy any algebraic constraint like positive definiteness. The 'completion problem' for partial correlation vines is therefore trivial. An incomplete specification of a partial correlation vine may be extended to a complete specification by assigning arbitrary numbers in the $(-1, 1)$ interval to the unspecified edges in the vine.

Partial correlation vines have another important property; the product of 1 minus the square partial correlations equals the determinant of the correlation matrix.

Theorem 3.2 *Let D be the determinant of the n-dimensional correlation matrix ($D > 0$). For any partial correlation vine*

$$D = \prod_{e \in E(\mathcal{V})} (1 - \rho^2_{e_1,e_2;D_e}) \tag{4}$$

Theorem 3.3 *Let g be an $n-$dimensional density satisfying the bivariate vine specification $(F, \mathcal{V}, B)$ with density g and one dimensional marginal densities $g_1, ..., g_n$; then*

$$I(g \mid \prod_{i=1}^{n} g_i) = \sum_{e \in E(\mathcal{V})} E_{D_e} I(g_{e_1,e_2|D_e} \mid g_{e_1|D_e} \cdot g_{e_2|D_e}). \tag{5}$$

If D_e is vacuous then by definition

$$E_{D_e} I(g_{e_1,e_2|D_e} \mid g_{e_1|D_e} \cdot g_{e_2|D_e}) = I(g_{e_1,e_2} \mid g_{e_1} \cdot g_{e_2})$$

Theorem 3.4 *Let $\mathcal{V} = (T_1, ..., T_{n-1})$ be a regular vine on n elements. For each edge $e \in E(\mathcal{V})$, let the conditional copula and copula density be $C_{e_1,e_2|D_e}$ and $c_{e_1,e_2|D_e}$. Let the marginal distributions F_i with densities $f_i, i = 1, ..., n$ be given. Then the vine-dependent distribution is uniquely determined and has a density given by*

$$f_{1\dots n} \quad = \quad f_1 \dots f_n \prod_{e \in E(\mathcal{V})} c_{e_1,e_2|D_e}(F_{e_1|D_e}, F_{e_2|D_e}). \tag{6}$$

4 Sampling

We assume that variables $X_1, X_2, ..., X_n$ are uniform on $(0, 1)$. Each edge in a regular vine may be associated with a conditional copula, that is, a conditional bivariate distribution with uniform margins . It is convenient to specify the conditional bivariate copulae by first assigning a constant conditional rank correlation to each edge of the vine. For $i = 1, \dots, n-1$, with $e \in E_i$ and $\{j, k\}$ the conditioned variables of e, D_e the conditioning variables of e, we associate

$$r_{j,k|D_e}.$$

The resulting structure is called a *conditional rank correlation vine*. Given a conditional rank correlation vine, we choose a class of copulae indexed by correlation coefficients in the interval $[-1, 1]$ and select the copulae with correlation corresponding to the conditional rank correlation assigned to the edge of the vine. A joint distribution satisfying the vine-copula specification can be constructed and sampled on the fly, and will preserve maximum entropy properties of the conditional bivariate distributions [3, 1].

The conditional rank correlation vine plus copula determines the whole joint distribution. There are two strategies for sampling such a distribution, which we term the *cumulative* and *density* approaches.

D-vine We first illustrate the cumulative approach with the distribution specified by the D-vine in Figure 1, D(1,2,3,4): Sample four independent variables distributed uniformly on interval [0,1], U_1, U_2, U_3, U_4 and calculate values of correlated variables X_1, X_2, X_3, X_4 as follows:

1. $x_1 = u_1$,
2. $x_2 = F^{-1}_{r_{12};x_1}(u_2)$,
3. $x_3 = F^{-1}_{r_{23};x_2}\left(F^{-1}_{r_{13|2};F_{r_{12};x_2}(x_1)}(u_3)\right)$,
4. $x_4 = F^{-1}_{r_{34};x_3}\left(F^{-1}_{r_{24|3};F_{r_{23};x_3}(x_2)}\left(F^{-1}_{r_{14|23};F_{r_{13|2};F_{r_{23};x_2}(x_3)}(F_{r_{12};x_2}(x_1))}(u_4)\right)\right)$

where $F_{r_{ij|k};X_i}(X_j)$ denotes the cumulative distribution function for X_j, applied to X_j, given X_i under the conditional copula with correlation $r_{ij|k}$. Notice that the D-vine sampling procedure uses conditional and inverse conditional distribution functions. A more general form of the above procedure simply refers to conditional cumulative distribution functions:

$$\begin{aligned} x_1 &= u_1 \\ x_2 &= F^{-1}_{2|1:x_1}(u_2), \\ x_3 &= F^{-1}_{3|2:x_2}\left(F^{-1}_{3|12:F_{1|2}(x_1)}(u_3)\right), \\ x_4 &= F^{-1}_{4|3:x_3}\left(F^{-1}_{4|23:F_{2|3}(x_2)}\left(F^{-1}_{4|123:F_{1|23}(x_1)}(u_4)\right)\right). \end{aligned} \tag{7}$$

Figure 2 depicts the sampling of X_4 in the D-vine in Figure 1 with a "staircase graph". Following the dotted arrows, we start by sampling U_4 (realization u_4) and use this with the copula for the conditional rank correlation of $\{1, 4\}$ given $\{2, 3\}$ to find the argument of $F^{-1}_{4|23}$, etc. Notice that

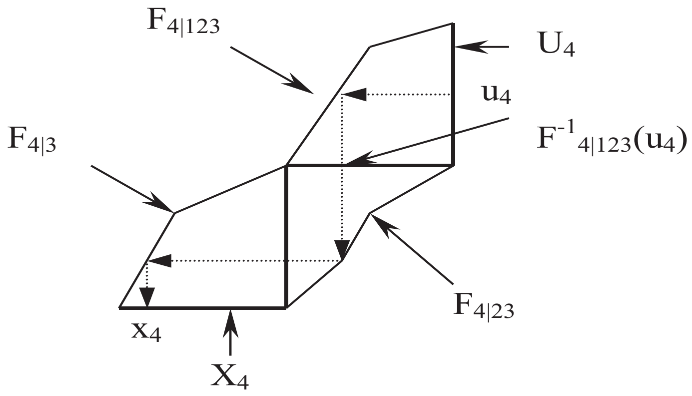

Figure 2: Staircase graph representation of D-Vine sampling procedure.

for the D-vine, values of $F_{2|3}$ and $F_{1|23}$ that are used to conditionalize copulae with correlations $r_{24|3}$ and $r_{14|23}$ to obtain $F_{4|23}$ and $F_{4|123}$, respectively, have to be calculated.

The staircase graph shows that if any of the cumulative conditional distributions in Figure 2 is uniform, then the corresponding abscissa and ordinates can be identified. This corresponds to noting that the inverse cumulative function in (7) is the identity, and this in turn corresponds to a conditional rank correlation being zero and the corresponding variables being conditionally independent. Notice that the conditional rank correlations can be chosen arbitrarily in the interval $[-1, 1]$; they need not be positive definite or satisfy any further algebraic constraint.

4.1 Sampling an arbitrary Regular Vine

A regular vine on n nodes will have a single node in tree $n-1$. It suffices to show how to sample one of the conditioned variables in this node, say n. Assuming we have sampled all the other variables we proceed as follows:

1. By Lemma 3.2, the variable n occurs in trees $1, ..., n-1$ exactly once as a conditioned variable. The variable with which it is conditioned in tree j is called its "j-partner". We define an ordering for n as follows: index the j-partner of variable n as variable j. We denote the conditional bivariate constraints corresponding to the partners of n as:

$$(n, 1|\emptyset), (n, 2|D_2^n), (n, 3|D_3^n)...(n, n-1|D_{n-1}^n)$$

   Again by Lemma 3.2, variables $1, ..., n-1$ appear first as conditioned variables (to the left of "|") before appearing as conditioning variables (to the right of "|"). Also,

$$0 = \#D_1^n < \#D_2^n < \ldots < \#D_{n-1}^n = n-2.$$

2. Assuming we have sampled all variables except n, sample one variable uniformly distributed on the interval (0,1), denoted u_n. We use the general notation $F_{a|b,C}$ to denote $F_{a,b|C:F_{b|C}}$; that is the conditional copula for $\{a, b|C\}$ conditional on a value of the cumulative conditional distribution $F_{b|C}$. Here, $\{a, b|C\}$ is the conditional bivariate constraint corresponding to a node in the vine.

3. Sample x_n as follows:

$$x_n = F^{-1}_{n|1,D_1^n}\left(F^{-1}_{n|2,D_2^n}\left(...\left(F^{-1}_{n|n-1,D_{n-1}^n}(u_n)\right)...\right)\right). \qquad (8)$$

The innermost term of (8) is:

$$\begin{aligned} F^{-1}_{n|n-1,D^n_{n-1}} &= F^{-1}_{n,n-1|D^n_{n-1}:F_{n-1|D^n_{n-1}}} \\ &= F^{-1}_{n,n-1|D^n_{n-1}:F_{n-1,n-2|D^{n-1}_{n-2}:F_{n-2|D^{n-1}_{n-2}}}} \end{aligned}$$

Density approach sampling When the vine-copula distribution is given as a density, the density approach to sampling may be used. If E is the edge set for $\mathcal{V}(n)$, for $e \in E$ with conditioning set D_e, let $c_{ij|D_e}$ be the copula density associated with e, then the density for a distribution specified by the assignment of copulae to the edges of $\mathcal{V}(n)$ is given by [1]:

$$\prod_{e\in E} c_{ij|D_e}(F_{i|D_e}(x_i), F_{j|D_e}(x_j)) f_1(x_1) \dots f_n(x_n) = \prod_{e\in E} c_{ij|D_e}(F_{i|D_e}(x_i), F_{j|D_e}(x_j)) \tag{9}$$

where, by uniformity, the density $f_i(x_i) = 1$.

This expression may be used to sample the vine distribution; namely, draw a large number of samples $(x_1, \dots, x_n)$ uniformly, and then resample these with probability proportional to (9). This is less efficient than the general sampling algorithm given previously; however it may be more convenient for conditionalization.

5 Continuous Bayesian Belief Nets

Bayesian belief nets are directed acyclic graphs that together with conditional probability functions represent high dimensional uncertainty distributions [21, 7, 4, 8]. The nodes represent variables, which can be discrete or continuous, and the arcs represent causal/influential or functional relationships between variables.

Continuous Bbns [21, 23] developed for joint normal variables interpret "influence" of the parents on a child as partial regression coefficients when the child is regressed on the parents. They require means, conditional variances and partial regression coefficients which can be specified in an algebraically independent manner. The restriction to joint normal cannot be easily relaxed. One cannot simply invoke the theory of linear least squares predictors as applied to arbitrary joint distributions. Suppose $(X_1, ..., X_{k-1})$ are the ancestors of X_k in an ordered Bbn[3]. We could interpret the "influence" of X_j on X_k as the partial regression of X_k on X_j given $1, ..., j-1, j+1, ..., k-1$. If j is not a parent of k, then j and k are conditionally independent given the parents of k; however, it is *not* the case that the partial regression of k on j, given the parents, is necessarily zero [13]. This means that the partial regression coefficients do not reflect the conditional independence structure of the Bbn.

[16] advances a vine-based distribution-free approach to continuous *Bbns*. Starting with an arbitrary Bbn whose nodes have continuous invertible distributions, we associate each arc with a (conditional) parent-child rank correlation according to a certain protocol. We specify nested sets of high dimensional joint distributions using the vine-copula approach, where any copula with invertible conditional cumulative distribution functions may be used so long as the chosen copula represents (conditional) independence as zero (conditional) correlation. The conditional rank correlations (like the partial regression coefficients) are algebraically independent, and there are tested protocols for their use in structured expert judgement [5, 12]. The vine generates a sampling algorithm which satisfies the conditional independence relations implied by the Bbn.

[3] Y is an ancestor of X with respect to an ordering of the variables which preserves the parent-child relations, that is, an ordering such that parents occur before their children in the ordering

The vine based approach is quite general, and of course this comes at a price: These Bbn's must be evaluated by Monte Carlo simulation. However, if the joint normal copula is used, then updating and conditionalizing can be done analytically [15, 6]

We associate nodes of a Bbn with univariate random variables $\{1, ..., n\}$ having uniform distributions on $(0, 1)$. We will associate the arcs, or "influences", with (conditional) rank correlations according to the following protocol:

1. 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. Index the nodes according to the sampling order $1, \ldots, n$.

2. Factorize the joint in the standard way following the sampling order. With sampling order $1, 2, \ldots, n$, write:

$$P(1, \ldots, n) \quad = \quad P(1)P(2|1)P(3|21) \ldots P(n|n-1, n-2, \ldots, 1).$$

3. Underscore those nodes in each condition, which are not parents of the conditioned variable and thus are not necessary in sampling the conditioned variable. This uses some of the conditional independence relations in the belief net. Hence if in sampling $2, \ldots, n$ variable 1 is not necessary (i.e. there is no influence from 1 to any other variable) then

$$P(1, \ldots, n) \quad = \quad P(1)P(2|\underline{1})P(3|2\underline{1}) \ldots P(n|n-1, n-2, \ldots, \underline{1}). \tag{10}$$

   The underscored nodes could be omitted thereby yielding the familiar factorization of the Bbn as a product of conditional probabilities, with each node conditionalized on its parents (for source nodes the set of parents is empty).

4. For each term i with parents (non-underscored variables) $i_1 ... i_{p(i)}$ in (10), associate the arc $i_{p(i)-k} \longrightarrow i$ with the conditional rank correlation

$$r(i, i_{p(i)}) \; ; k = 0$$

$$r(i, i_{p(i)-k} | i_{p(i)}, ..., i_{p(i)-k+1}); \; 1 \leq k \leq p(i) - 1. \tag{11}$$

   where the assignment is vacuous if $\{i_1 ... i_{p(i)}\} = \emptyset$. Assigning conditional rank correlations for $i = 1, ..n$, every arc in the Bbn is assigned a conditional rank correlation between parent and child.

Let $\mathcal{D}^i$ denote a D-vine on variables $1, ..., i$. The following theorem [16, 6] shows that these assignments uniquely determine the joint distribution and are algebraically independent:

Theorem 5.1 *Given a Bbn with n nodes and continuous invertible univariate margins; the specification of conditional rank correlations (11), $i = 1, ..., n$ and a copula realizing all correlations $[-1, 1]$ for which correlation 0 entails independence uniquely determines the joint distribution. This joint distribution satisfies the characteristic factorization (10) and the conditional rank correlations in (11) are algebraically independent.*

6 Model Learning

An approach to model learning inspired by [24] was developed in [16], based on the factorization of the determinant in Theorem 3.2. We sketch here a more general approach based on the mutual information. Following [9, 10], the mutual information is taken as a general measure of dependence. The strategy is to choose a regular vine which captures the mutual information in a small number of conditional bivariate terms, and to find a copula which renders these mutual information values.

Theorem 6.1 *Let f be a joint normal density with mean vector zero, then*

$$MI(f) = -\frac{1}{2}\, ln(D)$$

Where D is the determinant of the correlation matrix.

Proof: From ([24] , prop.6.4.5,p 168).

Remark 6.1 *For a bivariate normal, theorem 6.1 says that $MI(f) = -(1/2)ln(1-\rho^2)$. Substituting the appropriate conditional bivariate normal distributions in the right hand side of (5) we find $MI(f) = -1/2\sum_{e\in E(\mathcal{V})} ln(1-\rho^2_{e_1,e_2;D_e})$, which agrees with Theorem 3.2.*

The determinant of a correlation matrix indicates the 'amount of linearity' in a joint distribution. It takes the value 1 if the variables are uncorrelated, and the value zero if there is a linear dependence. Theorem 6.1 suggest that

$$e^{-2MI(f)}$$

is the appropriate generalization of the determinant to capture general dependence.

Proposition 6.1 *$e^{-2MI(f)} = 1$ if and only if $f = \Pi f_i$ and $e^{-2MI(f)} = 0$ if and only if f has positive mass on a set of Πf_i measure zero.*

The multiple correlation $R_{1\{2,...n\}}$ is the correlation between X_1 and the best linear predictor of X_1 from $X_2, ...X_n$. The appropriate concept for generalized dependence might be called multiple information

Definition 6.1 (Multiple information) *The* **multiple information** *of X_1 with respect to $X_2, ...X_n$ is*

$$MI_{1\{2,...n\}} = I(f \mid f_1 f_{2,...n}) \tag{12}$$

The following theorem helps interpret this notion.

Theorem 6.2

1. *$MI_{1\{2,...n\}} = MI_{1\{\pi_2,...\pi_n\}}$ where π is any permutation of $\{2, ...n\}$;*
2. *$MI_{1\{2,...n\}} = I(f_{1,...n} \mid \Pi^n_{i=1} f_i) - I(f_{2,...n} \mid \Pi^n_{i=2} f_i)$*
3. *$MI_{1\{2,...n\}} = \sum_{e\in E(\mathcal{V});e_1=1} E_{D_e} I(f_{1,e_j|D_e} \mid f_{1|D_e} f_{e_j|D_e})$; where $\mathcal{V}$ is any regular vine on n elements.*
4. *$MI(f) = MI_{1\{2,...n\}} + MI_{2\{3,...n\}} + MI_{1\{n\}}$*

Proof: [1] is obvious. [2] is a calculation, [3] follows from Theorem 3.3, using a D-vine with 1 in the top node.

We now write Theorem 3.3 as

$$MI(f) = \sum_{\{i,j|K(ij)\}\in\mathcal{V}} b_{ij;K(ij)} \tag{13}$$

where $K(ij)$ is conditioning set for the node in $\mathcal{V}$ with conditioned set $\{i, j\}$. The terms $b_{ij;K(ij)}$ will depend on the regular vine which we choose to represent the second order structure, however the sum of these terms must satisfy (13). We seek a regular vine for which the terms $b_{ij;K(ij)}$ in (13) are "as spread out" as possible. In other words, we wish to capture the total dependence $MI(f)$ in a small number of terms, with the remaining terms being close to zero. This concept is made precise with the notion of *majorization* [20].

Definition 6.2 *Let* $x,y \in \mathbb{R}^n$ *be such that* $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$*; then* x *majorizes* y *if for all* $k; k = 1, ..., n$

$$\sum_{j=1}^{k} x_{(j)} \leq \sum_{j=1}^{k} y_{(j)} \tag{14}$$

where $x_{(j)}$ *is the increasing arrangement of the components of* x*, and similarly for* y*.*

In view of (13) the model inference problem may be cast as the problem of finding a regular vine whose terms $b_{ij;K(ij)}$ are non-dominated in the sense of majorization. In that case, setting the smallest mutual informations equal to zero will change the overall mutual information as little as possible. Pairs of variables whose (conditional) mutual information is zero, are (conditionally) independent. Finding non-dominated solutions may be difficult, but a necessary condition for non-dominance can be found by maximizing any Schur convex function.

Definition 6.3 *A function* $f : \mathbb{R}^k \to \mathbb{R}$ *is Schur convex if* $f(x) \geq f(y)$ *whenever* x *majorizes* y*.*

Schur convex functions have been studied extensively. A sufficient condition for Schur convexity is given by [20].

Proposition 6.2 *If* $f : \mathbb{R}^k \to \mathbb{R}$ *may be written as* $f(x) = \sum f_i(x_i)$ *with* f_i *convex, then* f *is Schur convex.*

Vine Inference Strategy

The following strategy for model inference suggests itself:

1. Choose a Schur convex function $f : \mathbb{R}^{\frac{n(n-1)}{2}} \to \mathbb{R}$;
2. Find a regular vine $\mathcal{V}(n)$ whose vector $b_{ij;K(ij)}$ maximizes f;
3. Set the mutual informations in $\mathcal{V}(n)$ equal to zero for which the terms $b_{ij;K(ij)}$ are smallest;
4. Associate copulae with the nodes in the vine, such that the non-zero mutual information values are preserved.

This approach to model learning should be viewed as a research program rather than as an applicable technique. Estimating mutual information, searching the set of regular vines, choosing a Schur convex function and choosing a copula on the basis of mutual information are subjects that require much future work.

7 Example

We illustrate features of dependence modelling with a simple example involving return on investment over 5 years. We invest equal amounts of money in stocks, bonds and real estate. The yearly rates of return are uncertain; the return rates within each sector are correlated, and the sectors themselves are correlated. We assume that the bond market is negatively correlated to the stock market (rank correlation -0.8) and positively correlated to the real estate market (rank correlation 0.5). Within each market, the successive yearly rates of return may be positively (stocks and bonds) or negatively correlated (real estate); however non-adjacent years have reverse correlations. Each of the markets is assigned a different copula, as shown in Figure (4). Starting with $\$100,000$ The fortune after 5years is

$$\text{Fortune} = 100,000[(1 + s_1)\dots(1 + s_5) + (1 + b_1)\dots(1 + b_5) + (1 + r_1)\dots(1 + r_5)]$$

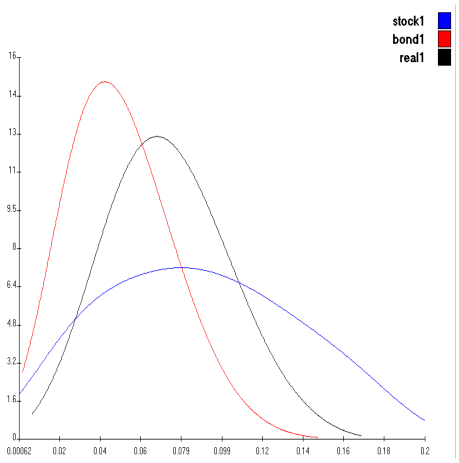

Figure 3: Univariate beta densities for stocks, bonds and real estate.

where s_1 is the return for the stocks in year 1, etc. The univariate densities are shown in Figure 3. The dependence structure is a tree connecting stocks, bonds and real estate, where each of the latter are represented as a vine. This is shown in Figure 4.

In spite of the symmetry of the fortune function, the complex dependence structure causes the variables to have different relations to the fortune. We see that variable $stock4$ has the highest correlation with Fortune. The regression on some of the variables is non linear (Figure 5).

The entire joint distribution can be viewed as a cobweb plot. Figure 6 shows a cobweb plot in which we have conditionalized on high and low values of Fortune. The strong relation with the stock returns in years 3 and 4 is clearly visible. The distribution can also be modelled as a Bbn. Figure 7 shows this, where we have conditionalized on a low return for stocks in year 1. The prior marginal distributions are shown in grey, the updated distributions in black.

```
Dependence Tree
Bivariate type: Elliptical ... Tree name: st_bn_rl
 stocks
    └(-0.80) bonds
      └(0.50) realestate

D-Vine
Copula type: Elliptical ... Vine name: stocks
stock1    stock2    stock3    stock4    stock5
 └ 0.80┘ └ 0.80┘ └ 0.80┘ └ 0.80┘
     └ -0.80┘ └ -0.80┘ └ -0.80┘
            └ 0.80┘ └ 0.80┘
                 └ 0.00┘

Canonical Vine
Copula type: Diagonal Band ... Vine name: bonds
bond1
├(0.80) bond2  bond2|bond1
├(0.80) bond3  ├(-0.80) bond3  bond3|bond2,bond1
├(0.80) bond4  ├(-0.80) bond4  ├(0.80) bond4  bond4|bond3,bond2,bond1
└(0.80) bond5  └(-0.80) bond5  └(0.80) bond5  └(0.00) bond5

D-Vine
Copula type: Normal ... Vine name: realestate
real1     real2      real3      real4     real5
 └ -0.80┘ └ -0.80┘ └ -0.80┘ └ -0.80┘
      └ 0.80┘     └ 0.80┘    └ 0.80┘
           └ -0.80┘ └ -0.80┘
                 └ 0.80┘
```

Figure 4: Tree with three vines for stocks, bonds and real estate.

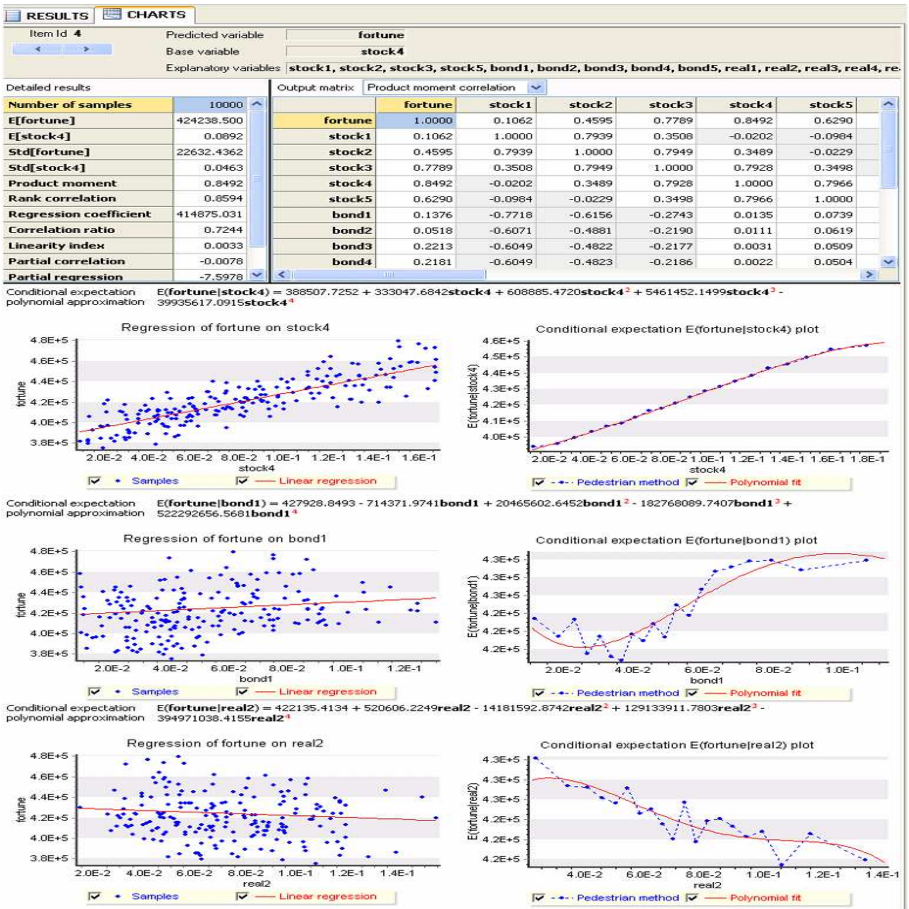

Figure 5: Conditional expectation

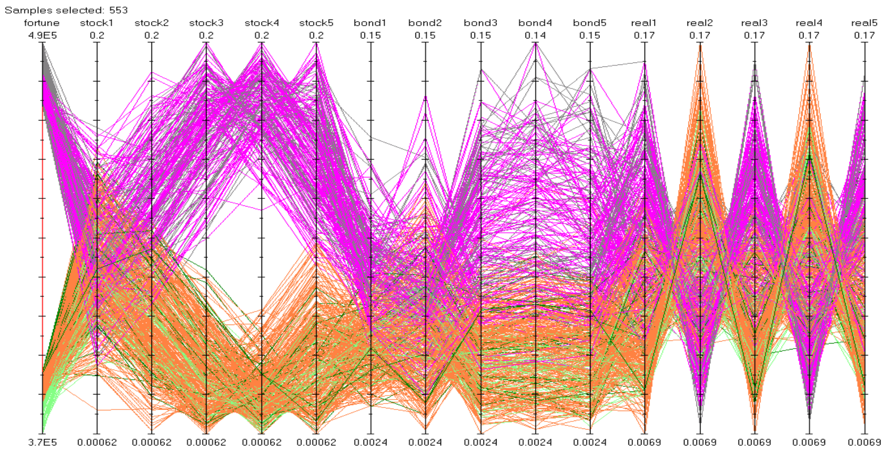

Figure 6: Cobweb plot, conditional on high and low values of Fortune

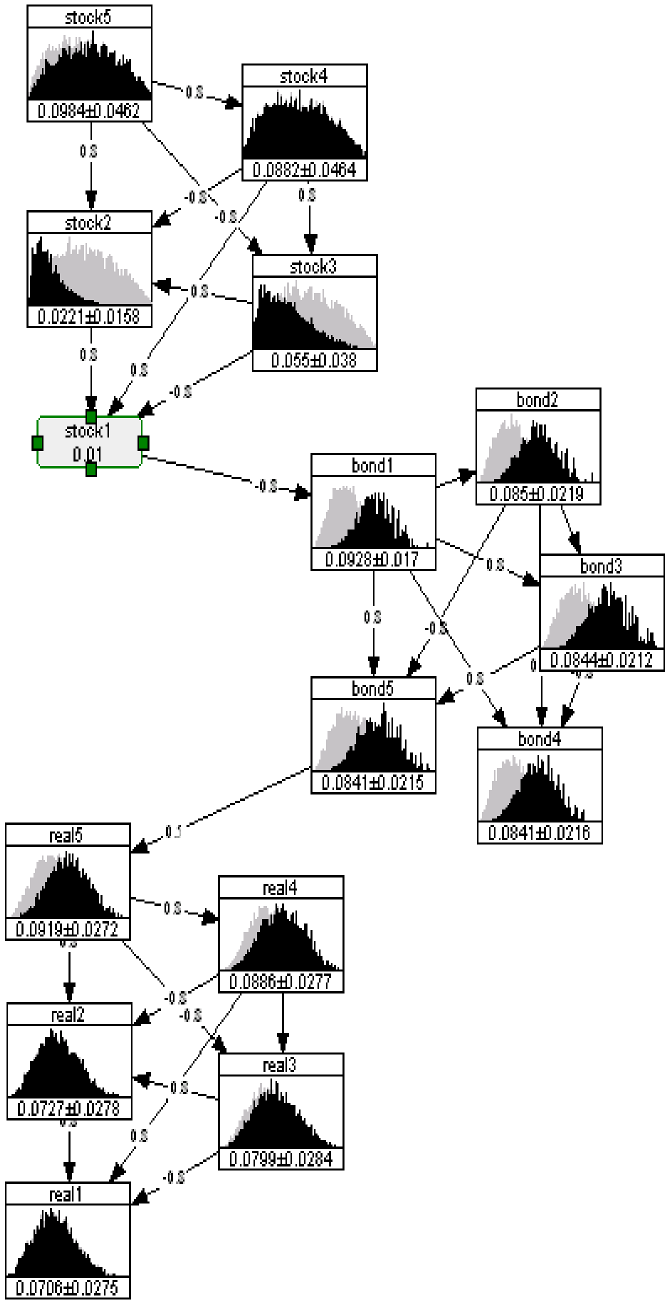

Figure 7: Bbn representation of Stocks, Bonds and Real estate, conditionalized on low value of stocks in year 1

8 Conclusions

This paper has illustrated the modelling resources and flexibility afforded by regular vines. Applications to date have been in uncertainty analysis, where, as in the example in section 7, we explore the effects of different dependence structures. Applications to model learning are a topic for the future. This would require a technique for estimating mutual correlation, a method for searching vines and associating copulae with conditional bivariate distributions on the basis of mutual information.

References

[1] T.J. Bedford and R.M Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268, 2001.

[2] T.J. Bedford and R.M. Cooke. Vines - a new graphical model for dependent random variables. *Ann. of Stat.*, 30(4):1031–1068, 2002.

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

[4] R.G. Cowell, A.P Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Statistics for Engineering and Information Sciences. Springer- Verlag, New York, 1999.

[5] L.H.J. Goossens, J.D. Harrison, B.C.P. Kraan, R.M. Cooke, F.T. Harper, and S.C. Hora. Probabilistic accident consequence uncertainty analysis: uncertainty assessment for internal dosimetry. Technical report, NUREG/CR-6571, EUR 16773, SAND98-0119, vol. 1, 1997.

[6] D. Hanea, A. Kurowicka and R.M. Cooke. Hybrid method for quantifying and analyzing bayesian belief nets. *Quality and Reliability Engineering International*, 22:709–729, 2006.

[7] F.V. Jensen. *An introduction to Bayesian Networks.* Taylor and Francis, London, United Kingdom, 1996.

[8] F.V. Jensen. *Bayesian Networks and Decision Graphs.* Springer-Verlag, New York, 2001.

[9] H. Joe. Multivariate dependence measures and data analysis. *Comp. Stat. and Data Analysis*, 16:279–297, 1993.

[10] H. Joe. Relative entropy measures of multivariate dependence. *J. Amer. Strat. Assoc*, 84(405):157–164, 1999.

[11] M.G. Kendall and A. Stuart. *The advenced theory of statistics.* Charles Griffin & Company Limited, London, 1961.

[12] B.C.P. Kraan. *Probabilistic Inversion in Uncertainty Analysis and related topics.* Isbn 90-9015710-7, PhD dissertation, TU Delft, 2002.

[13] D. Kurowicka and R.M. Cooke. Conditional and partial correlation for graphical uncertainty models. *in Recent Advances in Reliability Theory, Birkhauser, Boston*, pages 259–276, 2000.

[14] D. Kurowicka and R.M. Cooke. A parametrization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications*, 372:225–251, 2003.

[15] D. Kurowicka and R.M. Cooke. Distrib ution-free continuous bayesian belief nets. In *Mathematical Methods in Reliability*, 2004.

[16] D. Kurowicka and R.M. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling.* Wiley, 2006.

[17] D Kurowicka, R.M. Cooke, and U Callies. Vines inference. Accepted.

[18] D. Kurowicka and Cooke R.M. Completion problem with partial correlation vines. *Linear Algebra and Its Applications*, 418(1):188–200, 2006.

[19] D. Kurowicka and Cooke R.M. Sampling algorithms for generating joint uniform distributions using the vine - copula method. *Computational Statistics and Data Analysis*, 2007. doi:10.1016/j.csda.2006.11.043.

[20] A.W. Marshall and I. Olkin. *Inequalities: theory of Majorization and its Applications.* Academic Press, San Diego, 1979.

[21] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman Publishers, San Mateo, 1988.

[22] M. Rosenblat. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 27:470–472, 1952.

[23] R.D. Shachter and C.R. Kenley. Gaussian influence diagrams. *Menagement Science*, 35(5):527–550, 1989.

[24] J. Whittaker. *Graphical Models in applied multivariate statistics.* John Wiley and Sons, Chichester, 1990.

[25] G.U. Yule and M.G. Kendall. *An introduction to the theory of statistics.* Charles Griffin & Co. 14th edition, Belmont, California, 1965.