Annals of Mathematics and Artificial Intelligence **32:** 245–268, 2001.
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Probability density decomposition for conditionally dependent random variables modeled by vines

Tim Bedford [a,b] and Roger M. Cooke [b]

[a] *Department of Management Science, Strathclyde University, 40 George St, Glasgow G1 1QE, Scotland*
E-mail: tim@mansci.strath.ac.uk

[b] *Faculty of Information Technology and Systems, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands*
E-mail: r.m.cooke@its.tudelft.nl

A vine is a new graphical model for dependent random variables. Vines generalize the Markov trees often used in modeling multivariate distributions. They differ from Markov trees and Bayesian belief nets in that the concept of conditional independence is weakened to allow for various forms of conditional dependence. A general formula for the density of a vine dependent distribution is derived. This generalizes the well-known density formula for belief nets based on the decomposition of belief nets into cliques. Furthermore, the formula allows a simple proof of the Information Decomposition Theorem for a regular vine. The problem of (conditional) sampling is discussed, and Gibbs sampling is proposed to carry out sampling from conditional vine dependent distributions. The so-called 'canonical vines' built on highest degree trees offer the most efficient structure for Gibbs sampling.

Keywords: correlation, dependence, information, multivariate probability distribution, Monte-Carlo simulation, tree dependence, vine dependence, Markov tree, Bayesian belief net, Gibbs sampling

AMS subject classification: 62E10, 62E25, 62H20, 62B10, 94A17

1. Introduction

Graphical dependency models such as Markov trees, Bayesian belief networks and influence diagrams have become very popular in recent years. The principal applications of these graphical models have been in problems of Bayesian inference (Markov trees and belief nets), and to decision problems (influence diagrams). Markov trees have also been used within the area of uncertainty analysis to build multivariate dependent distributions (see, e.g., [2,3,14,15] and also [8,9] for other approaches based on the multivariate normal distribution).

Within the field of uncertainty analysis, the problem of easily specifying a coupling between two groups of random variables is prominent. Typically, information about marginal distributions is given, for example quantile information from experts. Elicitation of a full set of marginals and conditional distributions is however too great a burden. The methods referenced above all use some from of expert assessment of unconditional correlations to give partial information about the dependence structure. The methods assume that the distributions are members of some particular distributional class that is parameterized by the unconditional correlations and marginal distributions. Under such an assumption the multivariate distribution is then identified by the expert information.

The Markov tree methods described above are suitable for rapid Monte-Carlo simulation, thus reducing the computational burden of sampling from a high dimensional distribution. The bivariate joint distributions required to determine such a model exactly are chosen to have minimum information with respect to the independent distribution with the same marginals, under the conditions of having the correct marginals and the given rank correlation specified by an expert.

However, the assumption of conditional independence required for the use of Markov trees is rather strong and prohibits the modelling of certain common kinds of phenomena. Similarly, the use of standard graphical models to simplify the decomposition of the full distribution is often impossible because the requirement of conditional independence is too strong.

In [1,4], we introduced the notion of a vine dependent distribution and demonstrated the existence of such distributions. A vine is a graphical tool that allows the conditional independence property used in Markov trees and belief nets to be weakened. A vine allows the expert to specify conditional rank correlations, or more generally conditional copulae, in such a way that the information is *never* inconsistent. A special case is the specification of a multivariate normal distribution using partial correlations, generalizing work of Joe [11].

To demonstrate the kind of problem that we want to deal with using vines, consider three random variables X_1, X_2 and X_3 with uniform marginal distributions. We want to specify the joint distribution in a simple way. Figure 1 shows examples of (a) a belief net, (b) a Markov tree, and (c) a vine on three elements. In the case of the belief net and the Markov tree, variables 1 and 3 are conditionally independent given variable 2. In the vine, in contrast, they are conditionally dependent, with a conditional correlation coefficient that depends on the value taken by variable 2. The belief net specification is difficult to do because it entails the input of conditional distributions and some marginals (for figure 1 this would be two conditional distributions and one marginal). The conditionals have to be chosen so that the remaining marginals are uniform – a task that is not easy. The Markov tree and vine models both allow the use of copulae and are better suited to building a model with given marginals. Using the Markov tree one could specify a rank correlation between variables 1 and 2, and between variables 2 and 3, and use the minimum information copulae with these rank correlations (this is discussed in

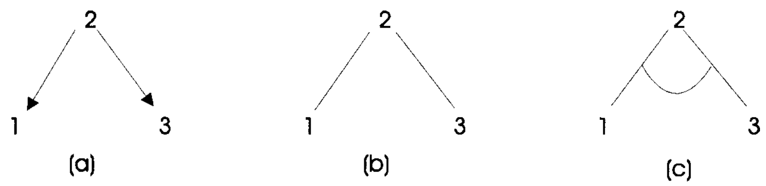
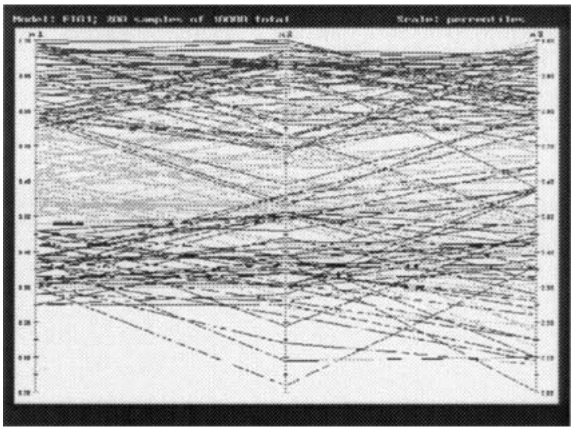
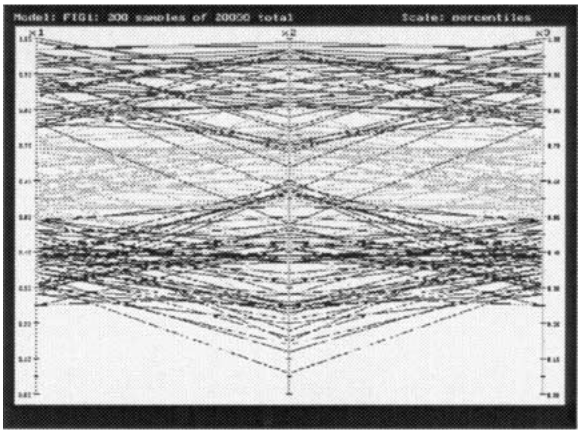


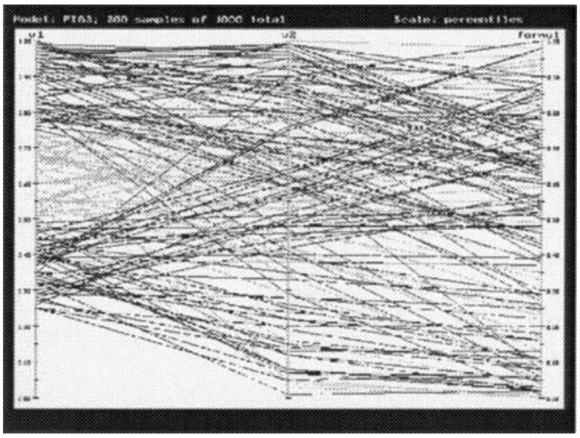
Figure 1. A belief net, a Markov tree, and a vine.



(a)



(b)



(c)

Figure 2. Plot with rank correlation of variables equal to (a) 0, (b) +1, and (c) −1.

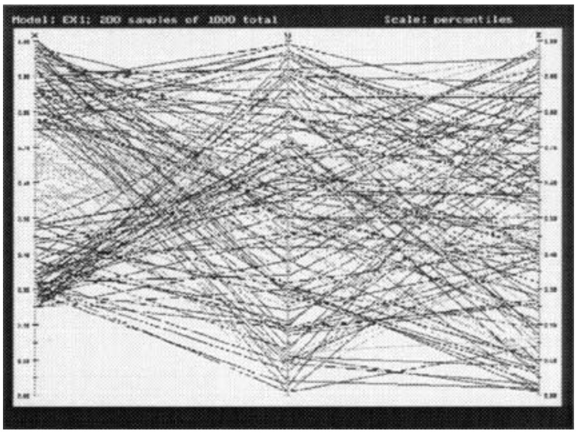
greater detail below). Suppose we were to take rank correlations of 0.9 in both cases. The Markov model is completely specified and gives variables 1 and 3 a rank correlation of approximately $0.9 \times 0.9 = 0.81$. However, we may not want to have this rank correlation! We may want a different rank correlation, or indeed to have the degree of correlation of variables 1 and 3 to depend on the value taken by variable 2. Figure 2 shows three plots. Each shows 200 independent samples of the three variables (the vertical lines are axes for each variable, and the lines correspond to samples). The first plot shows a sample obtained from a Markov tree in which variables 1 and 3 are independent given variable 2. The second plot shows a sample in which variables 1 and 3 have conditional rank correlation 1 given variable 2. The third plot shows a sample in which variables 1 and 3 have conditional rank correlation -1 given variable 2. Figure 3 shows three plots illustrating a distribution in which the conditional rank correlation of variables 1 and 3 given 2 is linearly dependent on variable 2 (the conditional rank correlation is $+1$ if $X_2 = 1$ and is -1 if $X_2 = 0$). Variables 1 and 2, and variables 2 and 3, have zero rank correlation. The first plot shows an unconditional sample. The second plot shows a sample conditioned on a high value for variable 2. The third plot shows a sample conditioned on a low value for variable 2. The kind of conditional dependence structure shown in figure 3 is difficult to model with a Markov tree (it is not impossible as there is an existence theorem that shows that *any* multivariate distribution on n variables can be modeled by a Markov tree with $n + 1$ variables – the proof is however highly non-constructive and thus has no practical significance!).

Vines were first defined in [4]. The existence of vine dependent distributions was demonstrated in [1], together with relative information properties, using a more general but non-graphical construction called a Cantor tree. Rank and partial correlation vines were studied extensively in [13], in particular, in the context of the problem of completing a positive definite matrix. While [1] makes very weak assumptions about the underlying distributions, we shall assume here that densities exist and concentrate on a simple subclass of vines called *regular vines*.

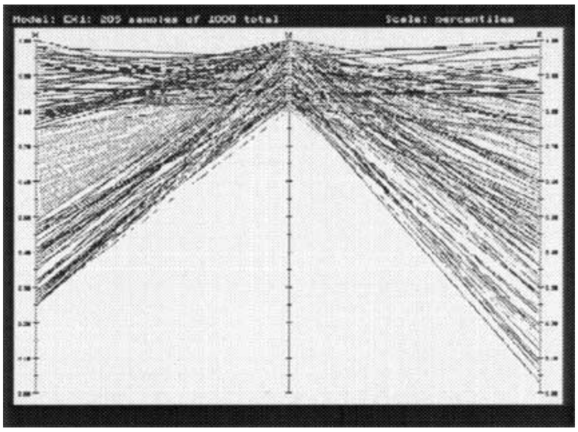
2. Definitions and preliminaries

We consider continuous probability distributions F on $\mathbb{R}^n$ equipped with the Borel sigma algebra $\mathcal{B}$. The one-dimensional marginal distribution functions of F are denoted F_i $(1 \leqslant i \leqslant n)$, the bivariate distribution functions are F_{ij} $(1 \leqslant i \neq j \leqslant n)$, and $F_{i|j}$ denotes the distribution of variable i conditional on j. The same subscript conventions apply to densities f and to higher-dimensional conditional marginals. For convenience we shall always assume that densities (with respect to Lebesgue measure) exist.

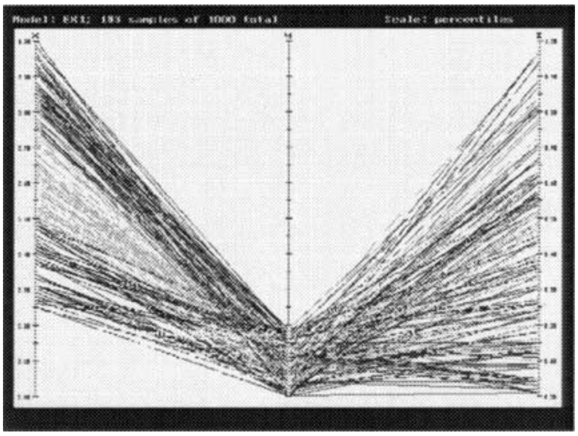
Definition 1 (Relative information). Let f and g be probability densities on $\mathbb{R}^n$ such that f is absolutely continuous with respect to g (that is, $g(x) = 0$ implies $f(x) = 0$), then the *relative information* or *Kullback–Liebler divergence*, $I(f \mid g)$ of f with respect



(a)



(b)



(c)

Figure 3. Plot showing the effect of conditional rank correlation.

to g is

$$I(f \mid g) = \int \log\left(\frac{f(x)}{g(x)}\right) f(x)\,\mathrm{d}x$$

(with the convention that $0/0 = 1$). When f is not absolutely continuous with respect to g we define $I(f \mid g) = \infty$.

The principle underlying partial specification of a probability distribution via (conditional) correlation coefficients and marginals is that the constructed distribution should be as "independent" as possible given the correlation constraints. Hence we will usually consider the relative information of a multivariate distribution with respect to the unique independent multivariate distribution having the same marginals.

Relative information $I(f \mid g)$ can be interpreted as measuring the lack of uniformity of f (with respect to g). The relative information is always non-negative and equals zero if and only if $f = g$ (see for example [12]).

Definition 2 (Rank or Spearman correlation). *The rank correlation* $r(X_1, X_2)$ of two random variables X_1 and X_2 with joint probability distribution F_{12} and marginal probability distributions F_1 and F_2 respectively, is given by

$$r(X_1, X_2) = \rho\big(F_1(X_1), F_2(X_2)\big).$$

Here $\rho(U, V)$ denotes the ordinary product moment correlation given by

$$\rho(U, V) = \frac{\operatorname{cov}(U, V)}{\sqrt{\operatorname{var}(U)\operatorname{var}(V)}},$$

and defined to be 0 if either U or V is constant. When Z is a random vector we can consider the conditional product moment correlation of U and V, $\rho_Z(U, V)$, which is simply the product moment correlation of the variables when conditioned on Z. The conditional rank correlation of X_1 and X_2 given Z is

$$r_Z(X_1, X_2) = r\big(\widetilde{X}_1, \widetilde{X}_2\big),$$

where $(\widetilde{X}_1, \widetilde{X}_2)$ has the distribution of (X_1, X_2) conditioned on Z.

The rank correlation has some important advantages over the ordinary product moment correlation:

- The rank correlation always exists.
- Independent of the marginal distributions F_X and F_Y it can take any value in the interval $[-1, 1]$ whereas the product moment correlation can only take values in a sub-interval $I \subset [-1, 1]$ where I depends on the marginal distributions F_X and F_Y.
- It is invariant under monotone increasing transformations of X and Y.

These properties make the rank correlation a suitable measure for developing canonical methods and techniques that are independent of marginal probability distributions.

The rank correlation is actually a measurement of the dependence of the *copula* between two random variables.

Definition 3 (Copula). The *copula* of two continuous random variables X and Y is the joint distribution of $(F_X(X), F_Y(Y))$.

Clearly, the copula of (X, Y) is a distribution on $[0, 1]^2$ with uniform marginals. More generally, we call any Borel probability measure μ a copula if $\mu([0, 1]^2) = 1$ and μ has uniform marginals.

An example of a copula is the *minimum information copula with given rank correlation*. This copula has minimum information with respect to the uniform distribution on the square, amongst all those copulae with the given rank correlation. The functional form of the density and an algorithm for approximating it arbitrarily closely are described in [16]. A second example is the *normal copula with correlation* ρ, obtained by taking (X, Y) to be joint normal with product moment correlation ρ in the definition of a copula given above. More information about copulae can be found in [17].

Definition 4 (Tree). A *tree* $T = \{N, E\}$ is an acyclic graph, where N is its set of nodes, and E is its set of edges (unordered pairs of nodes).

We begin by defining a tree structure that allows us to specify certain characteristics of a probability distribution.

Definition 5 (Bivariate tree specification). $(\underline{F}, T, B)$ is an n-dimensional *bivariate tree specification* if

(1) $\underline{F} = (F_1, \ldots, F_n)$ is a vector of one-dimensional distribution functions,

(2) T is a tree with nodes $N = \{1, \ldots, n\}$ and edges E,

(3) $B = \{B(i, j) \mid \{i, j\} \in E\}$, where $B(i, j)$ is a subset of the class of copula distribution functions.

Definition 6 (Tree dependence).

(1) A multivariate probability distribution G on $\mathbb{R}^n$ satisfies, or realizes, a bivariate tree specification $(\underline{F}, T, B)$ if the marginal distributions of G are F_i $(1 \leqslant i \leqslant n)$ and if for any $\{i, j\} \in E$ the bivariate copula C_{ij} of G is an element of $B(i, j)$;

(2) G has tree dependence of order M for T if $\{i, k_1\}, \ldots, \{k_m, j\} \in E$ implies that X_i and X_j are conditionally independent given any M of k_ℓ, $1 \leqslant \ell \leqslant m$; and if X_i and X_j are independent when there are no such $k_1, \ldots, k_m$ $(i, j \in N)$.

(3) G has Markov tree dependence for T if G has tree dependence order M for every $M > 0$.

In many applications it is convenient to take $B(i, j)$ to be the family of all copulae with a given rank correlation. This gives a rank correlation tree specification.

Definition 7 (Rank correlation tree specification). $(\underline{F}, T, t)$ is an n-dimensional rank correlation tree specification if

(1) $\underline{F} = (F_1, \ldots, F_n)$ is a vector of one-dimensional distribution functions,

(2) T is a tree with nodes $N = \{1, \ldots, n\}$ and edges E,

(3) The rank correlations of the bivariate distributions F_{ij}, $\{i, j\} \in E$, are specified by $t = \{t_{ij} \mid t_{ij} \in [-1, 1],\ \{i, j\} \in E,\ t_{ij} = t_{ji},\ t_{ii} = 1\}$.

The following two results were proved in [15]. The first is actually a special case of the well-known decomposition of a distribution modeled by a Bayesian belief net given by forming cliques (see [10]). The main result of this paper will be a generalization of this theorem to the case of vine-dependent distributions. For this reason we give a short proof.

Theorem 1. Let $(\underline{F}, T, B)$ be an n-dimensional bivariate tree specification that specifies the marginal densities f_i, $1 \leqslant i \leqslant n$, and the bivariate densities f_{ij}, $\{i, j\} \in E$, the set of edges of T. Then there is a unique density f on $\mathbb{R}^n$ with marginals $f_1, \ldots, f_n$ and bivariate marginals f_{ij} for $\{i, j\} \in E$ such that f has Markov tree dependence described by T. The density f is given by

$$f(x_1, \ldots, x_n) = \frac{\prod_{(i,j)\in E} f_{ij}(x_i, x_j)}{\prod_{i\in N} (f_i(x_i))^{\deg(i)-1}}, \tag{1}$$

where $\deg(i)$ denotes the degree of node i; that is, the number of neighbours of i in the tree T.

Sketch of proof. This is by induction on the number of nodes in the tree. For $n = 1, 2, 3$ the result is easy to show. Now assume that it holds for any tree with less than or equal to n nodes. Take a tree with $n + 1$ nodes. We can find a node (call it node $n + 1$) with degree 1, and thus attached by an edge to just one other node (call it node n). Now, recalling that $f_{1\ldots nn+1}$ is just the joint density of all $n + 1$ nodes, write

$$f_{1\ldots nn+1} := f_{n+1|n} f_{1\ldots n} = \frac{f_{nn+1}}{f_n} f_{1\ldots n}$$

and apply the formula for the subtree built from nodes $1, \ldots, n$, noting that node n has degree in this tree exactly one less than in the larger tree. This completes the proof. $\square$

The following theorem states that a rank correlation tree specification is always consistent. This gives part of the mathematical underpinning to the strategy of eliciting correlations from experts.

Theorem 2. Let $(\underline{F}, T, t)$ be an n-dimensional rank correlation tree specification, then there exists a joint probability distribution G realizing $(\underline{F}, T, t)$ with G Markov tree dependent.

Proof. This follows from theorem 1 if we can find appropriate bivariate distributions with the given marginals and rank correlations. Such bivariate distributions are given by applying the minimum information copula with given correlation $C_{\rho(i,j)}$,

$$F_{ij}(x_i, x_j) = C_{\rho(i,j)}\big(F_i(x_i), F_j(x_j)\big). \qquad \square$$

Theorem 2 would not hold if we replaced rank correlations with product moment correlations in definition 7. For arbitrary continuous and invertible one-dimensional distributions and an arbitrary $\rho \in [-1, 1]$, there need not exist a joint distribution having these one-dimensional distributions as marginals with product moment correlation ρ. (To see this recall that the product moment correlation of X and Y is equal to one if and only if $Y = aX + b$ for some $a > 0$. If this holds then the distribution functions are also related by $F_Y(at + b) = F_X(t)$, and so X is uniform if and only if Y is uniform. Hence when X is uniform and Y is not uniform, for example when Y is distributed as X^{10}, then the product moment distribution of X and Y must always be less than 1.)

The main result of this paper is to show that the above results can be generalized for vines. In particular, we shall show that for the vine dependent distributions to be defined below (the class of which includes Markov tree dependent distributions), a formula similar to that of theorem 1 holds. The density formula of theorem 1 is generalized in theorem 3 to include a product of terms explicitly representing the dependencies specified in the vine.

3. Regular vines

In this section we recall the notion of a regular vine as defined in [1].

Tree specifications are limited by the maximal number of edges in the tree. For trees with n nodes, there are at most $n - 1$ edges. This means we can constrain at most $n - 1$ bivariate marginals. By comparison there are $n(n-1)/2$ potentially distinct off-diagonal terms in a (rank) correlation matrix. The regular vine provides a more general structure for partially specifying joint distributions. For example, consider a density in three dimensions. In addition to specifying marginals g_1, g_2, and g_3, and rank correlations $r(X_1, X_2)$, $r(X_2, X_3)$, we also specify the conditional rank correlation of X_1 , and X_3 as a function of the value taken by X_2:

$$r_{x_2} = r(X_1, X_3 \mid X_2 = x_2).$$

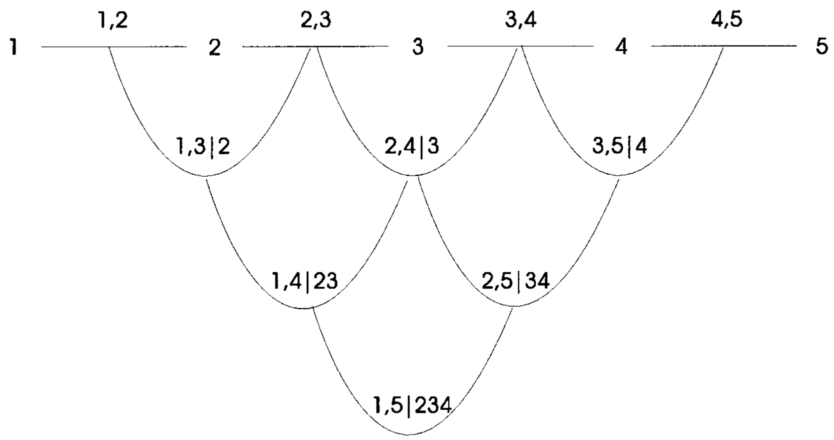


Figure 4. A regular vine.

For each value of X_2 we can specify a conditional rank correlation in $[-1, 1]$ and find the minimal information conditional distribution, provided the conditional marginals are not degenerate. This will be called a regular vine specification, and will be defined presently. Sampling such distributions on a computer is easily implemented; we simply use the minimal information distribution under a rank correlation constraint, but with the marginals conditional on X_2. Figures 4 and 6 show regular vine specifications on 5 variables. Figure 4 corresponds to the structure studied by Joe [11]. Each edge of a regular vine is associated with a restriction on the bivariate or conditional bivariate distribution shown adjacent to the edge. For example, in figure 4 the edge marked 1, 2 is associated with a specification of the bivariate distribution of variables 1 and 2. The edge marked 1, 3|2 is associated with bivariate distributions of 1 and 3 conditional on 2.

Note that the bottom level restrictions on the bivariate marginals form a tree T_1 with nodes $1, \ldots, 5$. The next level forms a tree T_2 whose nodes are the edges E_1 of T_1, and so on. There is no loss of generality in assuming that the edges E_i, $i = 1, \ldots, n-1$, have maximal cardinality $n - i$, as we may "remove" any edge by associating with it the vacuous restriction.

A regular vine can be used to define a class of distributions that form a sub-class of the Cantor tree distributions defined in [1]. The Cantor tree distributions can also be graphically represented by a more general object called a vine. A vine is used to place constraints on a multivariate distribution in a similar way to that in which directed acyclic graphs are used to constrain multivariate distributions in the theory of Bayesian belief nets.

Definition 8 (Regular vine, vine). $\mathcal{V}$ is a *vine* on n elements if

(1) $\mathcal{V} = (T_1, \ldots, T_m)$,

(2) T_1 is a tree with nodes $N_1 = \{1, \ldots, n\}$ and a set of edges denoted E_1,

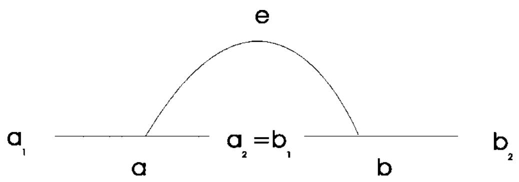


Figure 5. The proximity condition: an edge around a node.

(3) for $i = 2, \dots, m$, T_i is a tree with nodes $N_i \subset N_1 \cup E_1 \cup E_2 \cup \dots \cup E_{i-1}$ and edge set E_i.

A vine $\mathcal{V}$ is a *regular vine* on n elements if

(1) $m = n$,

(2) T_i is a connected tree with edge set E_i and node set $N_i = E_{i-1}$, with $\#N_i = n - (i-1)$ for $i = 1, \dots, n$, where $\#N_i$ is the cardinality of the set N_i,

(3) the proximity condition holds: for $i = 2, \dots, n-1$, if $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are two nodes in N_i connected by an edge $e \in E_i$ (recall $a_1, a_2, b_1, b_2 \in N_{i-1}$), then $\#a \cap b = 1$.

The proximity condition is illustrated in figure 5. In the situation shown there we say that edge e is *around* node a_2. It will be convenient to introduce some labeling corresponding to the edges and nodes in a vine, in order to specify the constraints. We first introduce a piece of notation to indicate which nodes of a tree with a lower index can be reached from a particular edge.

The edge set E_i consists of edges $e_i \in E_i$ which are themselves unordered pairs of nodes in N_i. Since $N_i \subset E_0 \cup E_1 \cup E_2 \cup \dots \cup E_{i-1}$ (where we write $N_1 = E_0$ for convenience), there exist $e_j \in E_j$ and $e_k \in E_k$ $(j, k < i)$ for which

$$e_i = \{e_j, e_k\}.$$

Definition 9. For any $e_i \in E_i$ the *complete union* of e_i is the subset

$$A_{e_i} = \bigl\{ j \in N_1 = E_0 \mid \exists 1 \leqslant i_1 \leqslant i_2 \leqslant \dots \leqslant i_r = i, \text{ and } e_{i_k} \in E_{i_k}\ (k = 1, \dots, r), \\ \text{with } j \in e_{i_1}, e_{i_k} \in e_{i_{k+1}}\ (k = 1, \dots, r-1) \bigr\}.$$

We can now define the constraint sets.

Definition 10 (Constraint set). For $e = \{e(1), e(2)\} \in E_\ell$, $\ell = 1, \dots, m-1$, the *conditioning set* associated with e is

$$D_e = A_{e(1)} \cap A_{e(2)},$$

and the *conditioned sets* associated with e are

$$C_{e,e(1)} = A_{e(1)} - D_e \quad \text{and} \quad C_{e,e(2)} = A_{e(2)} - D_e.$$

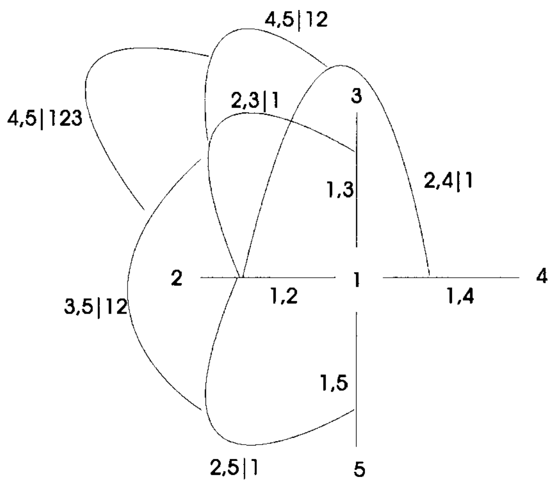


Figure 6. Canonical regular vine.

The *constraint set* for $\mathcal{V}$ is

$$\mathcal{CV} = \big\{(C_{e,e(1)}, C_{e,e(2)}, D_e) \mid \ell = 1, \dots, m-1,\ e \in E_\ell,\ e = \big\{e(1), e(2)\big\}\big\}.$$

Note that $A_e = A_{e(1)} \cup A_{e(2)} = C_{e,e(1)} \cup C_{e,e(2)} \cup D_e$ when $e = \{e(1), e(2)\}$. For $e \in E_m$ the conditioning set is empty.

The constraint set is shown for the regular vines in figures 4 and 6. At each edge $e \in E_\ell$, the terms $C_{e,e(1)}$ and $C_{e,e(2)}$ are separated by a comma and given to the left of the "|" sign, while D_e appears on the right. For example, in figure 4, the tree T_5 contains just a single node labeled $1, 5|234$. This node is the only edge of the tree T_4 where it joins the two (T_4)-nodes labeled $1, 4|23$ and $2, 5|34$. In this example the conditioned sets are always singletons. In fact this is always the case for regular vines.

The following result is proven in [1].

Lemma 1. If $\mathcal{V}$ is a regular vine on n elements then for all $\ell = 1, \dots, n-1$, and all $e \in E_\ell$ the conditioned sets associated with e are singletons, $\#C_{e,e(1)} = 1$. Furthermore, $\#A_e = \ell + 1$, and $\#D_e = \ell - 1$.

Because the conditioned sets are singletons, we shall often refer to the label of an edge e as being $i, j|D_e$. This is consistent with the labeling used in the figure, and simply means that the conditioned sets for e are $\{i\}$ and $\{j\}$. We will need the following proposition.

Proposition 1. Let $\mathcal{V}$ be a regular vine on n elements, and k an integer $1 \leqslant k < n-1$. Given a node i in tree T_k, there are exactly $\deg(i) - 1$ edges in T_{k+1} around i.

Proof. Without loss of generality we may assume that $k = 1$.

We show first that $\deg(i) - 1$ is the maximal number of edges in T_2 around i. Clearly there are $\deg(i)$ edges joining i to other nodes. These are the nodes in T_2 that when joined by edges will be around i. Because T_2 has to be a tree there can be no cycles of edges, so there are at most $\deg(i) - 1$ different edges in T_2 around i.

We now show that there are exactly $\deg(i) - 1$ edges in T_2 around i. Note first that any edge in T_2 can only be around one node of T_1 as otherwise there would be a cycle in T_1, and that by the proximity condition every edge in T_2 is around some node of T_1. If some node i of T_1 has less than $\deg(i) - 1$ edges in T_2 around it then we can count the total number of edges in T_2 as

$$\sum_{j \in T_1} \#\text{edges around } j < \sum_{j \in T_1} \bigl(\deg(j) - 1\bigr) = \left(\sum_{j \in T_1} \deg(j)\right) - n$$
$$= 2(n-1) - n = n - 2.$$

This contradicts the fact that T_2 has $n - 2$ edges. □

Using a regular vine we are able to partially specify a joint distribution as follows:

Definition 11 (Regular vine specification). $(\underline{F}, \mathcal{V}, B)$ is a *regular vine specification* if

(1) $\underline{F} = (F_1, \ldots, F_n)$ is a vector of continuous invertible distribution functions,

(2) $\mathcal{V}$ is a regular vine on n elements,

(3) $B = \{B_e(d) \mid i = 1, \ldots, n-1;\ e \in E_i\}$ where $B_e(d)$ is a collection of copulae and d is a vector of values taken by the variables in D_e.

The idea is that given the values taken by the variables in the constraint set D_e, the copula of the variables $X_{C_{e,j}}$ and $X_{C_{e,k}}$ must be a member of the specified collection of copulae.

Definition 12 (Regular vine dependence). A joint distribution F on variables $X_1, \ldots, X_n$ is said to *realize a regular vine specification* $(\underline{F}, \mathcal{V}, B)$ or exhibit *regular vine dependence* if for each $e \in E_i$, the copula of $X_{C_{e,j}}$ and $X_{C_{e,k}}$ given X_{D_e} is a member of $B_e(X_{D_e})$, and the marginal distribution of X_i is F_i $(i = 1, \ldots, n)$.

It is shown in [1] that, under the appropriate measurability conditions, regular vine dependent distributions can be constructed.

A convenient way to constrain the copulae in practice is to specify rank correlations and conditional rank correlations. In this case we talk about a *rank correlation vine specification*. Another way to constrain the copulae is by specifying a partial correlation. This is discussed in [1], where it is shown that a multivariate normal distribution can be conveniently parameterized by a partial correlation regular vine specification. The parameters of this specification are a set partial correlations without any algebraic relations. The advantage of this parameterization is that one does not have to worry about positive definiteness.

Example 1. We return to the example discussed above and illustrated by a sample in figure 3. This distribution is modeled by the vine shown in figure 1. The bivariate specification for variables X_1 and X_2 is that the rank correlation should be 0, as is the specification for variables X_2 and X_3. The conditional rank correlation of variables X_1 and X_3 given X_2 is equal to $2X_2 - 1$. For the simulation performed to produce figure 3 we used the minimum information distribution described in [1].

We now show that a formula for the density of a regular vine distribution can be given that is analogous to, and generalizes, the expression given in theorem 1 for Markov trees.

Lemma 2. Let $F_{12}(x, y)$ be a distribution function with topological support $B_1 \times B_2 \subset \mathbb{R}^2$ and density f_{12}. Suppose that the marginal densities f_1 and f_2 are strictly positive on B_1 and B_2, respectively. Let C be the copula of F_{12} and c its density. Then

$$c\big(F_1(x), F_2(y)\big) = \frac{f_{12}(x, y)}{f_1(x) f_2(y)}.$$

Proof. The copula is the push-forward distribution of F_{12} under the transformation

$$(x, y) \mapsto \big(F_1(x), F_2(y)\big).$$

The Jacobian of the transformation is $f_1(x) f_2(y)$ so the formula is just an application of the density transformation formula. □

Example 2. We begin by showing how the lemma can be applied to obtain the density of a vine dependent distribution for the regular vine in figure 4. The lemma is applied inductively to the edges of trees in reverse order, $T_4,\ T_3,\ T_2,\ T_1$. This calculation will be generalized in theorem 3.

$$\begin{aligned}
f_{12345} &= f_{15|234} f_{234} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) f_{1|234} f_{5|234} f_{234} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) \frac{f_{1234} f_{2345}}{f_{234}} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) \frac{f_{14|23} f_{23} f_{25|34} f_{34}}{f_{234}} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) c_{14|23}(F_{1|23}, F_{4|23}) c_{25|34}(F_{2|34}, F_{5|34}) \\
&\quad \times \frac{f_{1|23} f_{4|23} f_{23} f_{2|34} f_{5|34} f_{34}}{f_{234}} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) c_{14|23}(F_{1|23}, F_{4|23}) c_{25|34}(F_{2|34}, F_{5|34}) \frac{f_{123} f_{234} f_{345}}{f_{23} f_{34}} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) c_{14|23}(F_{1|23}, F_{4|23}) c_{25|34}(F_{2|34}, F_{5|34}) \frac{f_{13|2} f_{24|3} f_{35|4} f_2 f_3 f_4}{f_{23} f_{34}}
\end{aligned}$$

$$
\begin{aligned}
&= c_{15|234}(F_{1|234}, F_{5|234}) c_{14|23}(F_{1|23}, F_{4|23}) c_{25|34}(F_{2|34}, F_{5|34}) c_{13|2}(F_{1|2}, F_{3|2}) \\
&\quad \times c_{24|3}(F_{2|3}, F_{4|3}) c_{35|4}(F_{3|4}, F_{5|4}) \frac{f_{1|2} f_{3|2} f_{2|3} f_{4|3} f_{3|4} f_{5|4} f_2 f_3 f_4}{f_{23} f_{34}} \\
&= c_{15|234}(F_{1|234}, F_{5|234}) c_{14|23}(F_{1|23}, F_{4|23}) c_{25|34}(F_{2|34}, F_{5|34}) \\
&\quad \times c_{13|2}(F_{1|2}, F_{3|2}) c_{24|3}(F_{2|3}, F_{4|3}) c_{35|4}(F_{3|4}, F_{5|4}) \frac{f_{12} f_{23} f_{34} f_{45}}{f_2 f_3 f_4}.
\end{aligned}
$$

This is the form given in theorem 3 below and should be compared to the density for tree-dependent variables given in theorem 1. The final term is exactly what theorem 1 gives for the first level tree T_1 in the case of a Markov dependent distribution. Indeed, our formula specializes back to theorem 1 in this case, as it corresponds to using the independent copula everywhere, so that the (conditional) copula density terms are all identically 1.

In the next result we assume that, for each edge $e \in T_m$ ($m = 1, \ldots, n-1$), and each possible value of the variables in the conditioning set D_e a copula is specified. When the edge is labeled $ij|D_e$ we write the corresponding copula as $C_{ij|D_e}$ and its density as $c_{ij|D_e}$. The density of this copula is denoted c_d. Furthermore, marginal distributions F_j are specified for each $j \in N$.

Theorem 3. Let $\mathcal{V} = (T_1, \ldots, T_n)$ be a regular vine on n elements. Given F_i and $C_{ij|D_e}$ as above there is a unique vine dependent distribution with density given by

$$
f_{1\ldots n} = \left(\prod_{m=2}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}) \right) \frac{\prod_{(i,j) \in E_1} f_{ij}}{\prod_{i \in N_1} (f_i)^{\deg(i)-1}},
$$

where e is an edge with label $ij|D_e$.

Proof. The proof is by reverse induction on the level of the tree in the vine. We claim that for every $2 \leqslant M \leqslant n-1$,

$$
f_{1\ldots n} = \left(\prod_{m=M}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}) \right) \frac{\prod_{e \in E_{M-1}} f_{A_e}}{\prod_{e \in N_{M-1}} (f_{A_e})^{\deg(e)-1}}.
$$

The inductive claim certainly holds when $M = n-1$ as can be seen as follows. Writing $e = \{e_1, e_2\}$ for the one edge in T_{n-1}, with $A_{e_1} = \{i\} \cup D_e$ and $A_{e_2} = \{j\} \cup D_e$, we have

$$
\begin{aligned}
f_{1\ldots n} &= f_{ij|D_e} f_{D_e} = c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}) f_{i|D_e} f_{j|D_e} f_{D_e} \\
&= c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}) \frac{f_{A_{e_1}} f_{A_{e_2}}}{f_{D_e}}.
\end{aligned}
$$

Now since T_{n-2} is a tree with two edges and three nodes, one of the nodes, k say, must have degree two, and the edge e of T_{n-1} must be around k. Hence $D_e = A_k$, and the claim is demonstrated.

For the inductive step, assume that the formula holds for M. We show that it holds also for $M-1$. To see this, apply first the same decomposition as above for the marginal distribution corresponding to each edge of T_M. For $e \in E_{M-1}$ there are nodes of N_{M-1}, or, equivalently, edges of E_{M-2}, that we call $e(1)$ and $e(2)$ such that $e = \{e(1), e(2)\}$. The decomposition immediately gives all the claimed copula density terms, but the remaining term built from densities f is of the following form:

$$\left(\prod_{e \in E_{M-1}} \frac{f_{A_{e(1)}} f_{A_{e(2)}}}{f_{D_e}}\right) \frac{1}{\prod_{k \in N_{M-1}} (f_{A_k})^{\deg(k)-1}}.$$

In order to show that this reduces to the formula claimed for the induction step we have to show two things. Firstly, that the extra multiplicity of $f_{A_{e(1)}}$ terms arising because a node of N_{M-1} occurs in more than one edge of E_{M-1} is cancelled by $\prod_{k \in N_{M-1}} (f_{A_k})^{\deg(k)-1}$. Secondly, that

$$\prod_{e \in E_{M-1}} f_{D_e} = \prod_{k \in E_{M-2}} (f_{A_k})^{\deg(k)-1}.$$

The first claim is clear, since the degree of a node is just the number of edges it is in. Hence the multiplicity of a term $f_{A_{e(i)}}$ in the denominator is $\deg(e(i))$, so that after cancellation we retain each term exactly once.

For the second claim note that if $e \in E_{M-1}$ then D_e equals A_k for some $k \in N_{M-2}$, and furthermore that e is around k. The claim then follows immediately from proposition 1.

This completes the proof. □

By applying lemma 2 once more we can replace each term f_{ij} by $c_{\{i,j\}}(F_i, Fj) f_i f_j$. After cancelling terms we then obtain an alternative expression for the density.

Corollary 1. Let $\mathcal{V} = (T_1, \ldots, T_n)$ be a regular vine on n elements. Given F_i and $C_{ij|D_e}$ as above there is a unique vine dependent distribution with density given by

$$f_{1\ldots n} = f_1 \cdots f_n \prod_{m=1}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}),$$

where e is an edge labeled $ij|D_e$.

We now give another example of the decomposition to show that the tree structure can ensure a simple decomposition.

Example 3 (Canonical vine). We take the regular vine on 5 nodes shown in figure 6. Each tree here has a node with the highest possible degree. The node with highest degree has been chosen so that the conditioning sets are the same everywhere in each tree, and increase as {1}, {1, 2}, {1, 2, 3} as we move up through the trees. Such a vine is called a *canonical vine* in [13]. The calculation of the density is as follows:

$$
\begin{aligned}
f_{12345} &= f_{45|123} f_{123} \\
&= c_{45|123}(F_{4|123}, F_{5|123}) f_{4|123} f_{5|123} f_{123} \\
&= c_{45|123}(F_{4|123}, F_{5|123}) \frac{f_{34|12} f_{12} f_{35|12} f_{12}}{f_{123}} \\
&= c_{45|123}(F_{4|123}, F_{5|123}) c_{34|12}(F_{3|12}, F_{4|12}) c_{35|12}(F_{3|12}, F_{5|12}) \\
&\quad \times \frac{f_{3|12} f_{4|12} f_{12} f_{3|12} f_{5|12} f_{12}}{f_{123}} \\
&= c_{45|123}(F_{4|123}, F_{5|123}) c_{34|12}(F_{3|12}, F_{4|12}) c_{35|12}(F_{3|12}, F_{5|12}) \frac{f_{123} f_{124} f_{125}}{f_{12}^2} \\
&= c_{45|123}(F_{4|123}, F_{5|123}) c_{34|12}(F_{3|12}, F_{4|12}) c_{35|12}(F_{3|12}, F_{5|12}) c_{23|1}(F_{2|1}, F_{3|1}) \\
&\quad \times c_{24|1}(F_{2|1}, F_{4|1}) c_{25|1}(F_{2|1}, F_{5|1}) \frac{f_{2|1} f_{3|1} f_{2|1} f_{4|1} f_{2|1} f_{5|1} f_1^3}{f_{12}^2} \\
&= c_{45|123}(F_{4|123}, F_{5|123}) c_{34|12}(F_{3|12}, F_{4|12}) c_{35|12}(F_{3|12}, F_{5|12}) c_{23|1}(F_{2|1}, F_{3|1}) \\
&\quad \times c_{24|1}(F_{2|1}, F_{4|1}) c_{25|1}(F_{2|1}, F_{5|1}) \frac{f_{12} f_{13} f_{14} f_{15}}{f_1^3}.
\end{aligned}
$$

Applying lemma 2 again gives the expression of corollary 1:

$$
\begin{aligned}
f_{12345} &= c_{45|123}(F_{4|123}, F_{5|123}) c_{34|12}(F_{3|12}, F_{4|12}) c_{35|12}(F_{3|12}, F_{5|12}) c_{23|1}(F_{2|1}, F_{3|1}) \\
&\quad \times c_{24|1}(F_{2|1}, F_{4|1}) c_{25|1}(F_{2|1}, F_{5|1}) c_{12}(F_1, F_2) c_{13}(F_1, F_3) c_{14}(F_1, F_4) c_{15}(F_1, F_5) \\
&\quad \times f_1 f_2 f_3 f_4 f_5.
\end{aligned}
$$

Careful study of this expression shows that we can write the "standard" decomposition of the joint density,

$$
f_{12345} = f_1 \times f_{2|1} \times f_{3|21} \times f_{4|321} \times f_{5|4321}
$$

by

$$
\begin{aligned}
f_1 &= f_1, \\
f_{2|1} &= c_{12} f_2, \\
f_{3|21} &= c_{23|1} c_{13} f_3, \\
f_{4|321} &= c_{34|12} c_{24|1} c_{14} f_4, \\
f_{5|4321} &= c_{45|123} c_{35|12} c_{25|1} c_{15} f_5.
\end{aligned}
$$

No such simple expression exists for the standard vine.

4. Relative information of Markov tree dependent distributions

From theorem 1 it follows by a straightforward calculation that for the Markov tree dependent density g given by the theorem,

$$I\left(g \mid \prod_{i\in N} f_i\right) = \sum_{\{i,j\}\in E} I(f_{ij} \mid f_i f_j). \tag{2}$$

If the bivariate tree specification does not completely determine the bivariate marginals f_{ij}, $\{i, j\} \in E$, then more than one Markov tree dependent realization may be possible. In this case equation (2) shows that relative information with respect to the product distribution $\prod_{i\in N} f_i$ is minimized, within the class of Markov tree dependent realizations, by minimizing each bivariate relative information $I(f_{ij} \mid f_i f_j)$, $\{i, j\} \in E$.

Markov tree dependent distributions are optimal realizations of bivariate tree specifications in the sense of minimizing relative information with respect to the independent distribution with the same marginals. In other words, a minimal information realization of a (consistent) bivariate tree specification has Markov tree dependence. This follows from a very general result, proven in [1,4], stating that relative minimum information distributions (relative to independent distributions), subject to a marginal constraint on a subset of variables, have a conditional independence property given that subset:

Theorem 4. Assume that $g_{X,Y}$ is a probability density with marginals f_X and f_Y that uniquely minimizes $I(g_{X,Y} \mid f_X f_Y)$ within the class of distributions $B(X, Y)$. Assume similarly that $g_{X,Z}$ is a probability density with marginals f_X and f_Z that uniquely minimizes $I(g_{X,Z} \mid f_X f_Z)$ within the class of distributions $B(X, Z)$. Then $g_{X,Y,Z} := g_X g_{Y|X} g_{Z|X}$ is the unique probability density with marginals f_X, f_Y and f_Z that minimizes $I(g_{X,Y,Z} \mid f_X f_Y f_Z)$ with marginals $g_{X,Y}$ and $g_{X,Z}$ constrained to be members of $B(X, Y)$ and $B(X, Z)$, respectively.

Corollary 2. Let $(\underline{F}, T, B)$ be a bivariate tree specification. For each $(i, j) \in E$, let there be a unique density $g(x_i, x_j)$ which has minimum information relative to the product measure $f_i f_j$ under the constraint $B(i, j)$. Then the unique density with minimum information relative to the product density $\prod_{i\in N} f_i$ under constraints $B(i, j)$, $\{i, j\} \in E$ is obtained by taking the unique Markov tree dependent distribution with bivariate marginals $g(x_i, x_j)$, for each $\{i, j\} \in E$.

For regular vines it is possible to compute a useful expression for the information of a distribution in terms of the information of lower dimensional distributions. This Information Decomposition Theorem was already given for the more general case of a Cantor tree dependent distribution in [1], but with the result of theorem 3 we can give a much simpler proof for the regular vine case. In the proof we use the notation $\underline{x}$ to denote the full vector $(x_1, \ldots, x_n)$, and $\underline{x}_A$ to denote the subvector of those x_i with an index in A.

Theorem 5 (Information Decomposition Theorem). Let $\mathcal{V} = (T_1, \ldots, T_n)$ be a regular vine on n elements. For the unique vine dependent distribution with density given by

$$f_{1\ldots n} = f_1 \cdots f_n \prod_{m=1}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}),$$

where e is an edge labeled $ij|D_e$ the relative information of $f_{1\ldots n}$ with respect to the independent distribution with the same marginals is

$$I(f_{1\ldots n} \mid f_1 \cdots f_n) = \sum_{m=1}^{n-1} \sum_{e \in E_m} \mathbb{E}_{D_e} I(c_{ij|D_e}),$$

where $I(c)$ is the information of a copula with respect to the independent copula, and $\mathbb{E}_{D_e}$ denotes the expectation taken over the values of the variables in D_e.

Proof. We start with the form of the density given in corollary 1, namely,

$$f_{1\ldots n} = f_1 \cdots f_n \prod_{m=1}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}).$$

Hence

$$\begin{aligned}
I(f_{1\ldots n} \mid f_1 \cdots f_n) &= \int \log\left(\frac{f_{1\ldots n}(\underline{x})}{f_1(x_1) \cdots f_n(x_n)}\right) f_{1\ldots n}(\underline{x})\, \mathrm{d}x \\
&= \sum_{m=1}^{n-1} \sum_{e \in E_m} \int \log\bigl(c_{ij|D_e}(F_{i|D_e}, F_{j|D_e})\bigr) f_{1\ldots n}\, \mathrm{d}\underline{x} \\
&= \sum_{m=1}^{n-1} \sum_{e \in E_m} \int \log\bigl(c_{ij|D_e}(F_{i|D_e}, F_{j|D_e})\bigr) f_{A_e}\, \mathrm{d}\underline{x}_{A_e} \\
&= \sum_{m=1}^{n-1} \sum_{e \in E_m} \int \log\bigl(c_{ij|D_e}(F_{i|D_e}, F_{j|D_e})\bigr) f_{ij|D_e} f_{D_e}\, \mathrm{d}\underline{x}_{A_e} \\
&= \sum_{m=1}^{n-1} \sum_{e \in E_m} \mathbb{E}_{D_e}\Bigl(\log\bigl(c_{ij|D_e}(u, v) c_{ij|D_e}(u, v)\bigr)\, \mathrm{d}u\, \mathrm{d}v\Bigr) \\
&= \sum_{m=1}^{n-1} \sum_{e \in E_m} \mathbb{E}_{D_e}\bigl(I(c_{ij|D_e})\bigr). \qquad \square
\end{aligned}$$

As an example consider the regular vine shown in figure 4. We have,

$$\begin{aligned}
I(f_{12345} \mid f_1 \cdots f_5) = {} & I(f_{12} \mid f_1 f_2) + I(f_{23} \mid f_2 f_3) + I(f_{34} \mid f_3 f_4) + I(f_{45} \mid f_4 f_5) \\
& + \mathbb{E}_2 I(f_{13|2} \mid f_{1|2} f_{3|2}) + \mathbb{E}_3 I(f_{24|3} \mid f_{2|3} f_{4|3})
\end{aligned}$$

$$+ \mathbb{E}_4 I(f_{35|4} \mid f_{3|4} f_{5|4}) + \mathbb{E}_{23} I(f_{1,4|23} \mid f_{1|23} f_{5|23})$$
$$+ \mathbb{E}_{34} I(f_{2,5|34} \mid f_{2|34} f_{5|34}) + \mathbb{E}_{234} I(f_{1,5|234} \mid f_{1|234} f_{5|234}).$$

This expression shows that if we take a minimal information copula satisfying each of the (local) constraints, then the resulting joint distribution is also minimally informative.

Theorem 6. Let f be a Borel probability measure on $\mathbb{R}^n$ satisfying the regular vine specification $(\underline{F}, \mathcal{V}, B)$, and suppose that for each $e \in E_m$, with label $ij|D_e$ and for each value taken by the variables in D_e, $c_{ij|De}$ is a copula density minimizing

$$I(c_{ij|D_e}) \tag{3}$$

within the class of allowed copulae. Then f satisfies $(\underline{F}, \mathcal{V}, B)$ and minimizes

$$I\left(f \mid \prod_{i=1}^{n} f_i\right). \tag{4}$$

Furthermore, if any of the $c_{ij|D_e}$ uniquely minimizes the information term in (3) (for all values d of D_e), then f uniquely minimizes the information term in (4).

This result motivates the use of the bivariate minimum information distribution with given rank correlation.

5. Sampling and conditional sampling from vine distributions

Sampling from a regular vine distribution can be done in a straightforward way. We illustrate with the example in figure 4.

- Sample first X_1 according to the distribution function F_1.
- From C_{12}, F_1 and F_2 determine $F_{2|1}$ and sample X_2 given X_1.
- From C_{12}, F_1 and F_2 determine $F_{1|2}$. From C_{23}, F_2 and F_3 determine $F_{3|2}$. From $C_{13|2}$, $F_{1|2}$ and $F_{3|2}$ determine $F_{3|12}$ and sample X_3 given X_1 and X_2.
- From C_{34}, F_3 and F_4 determine $F_{4|3}$. From C_{23}, F_2 and F_3 determine $F_{2|3}$. From $C_{24|3}$, $F_{2|3}$ and $F_{4|3}$ determine $F_{4|23}$. From $C_{13|2}$, $F_{1|2}$ and $F_{3|2}$ determine $F_{1|23}$. From $C_{14|23}$, $F_{1|23}$ and $F_{4|23}$ determine $F_{4|123}$ and sample X_4 given X_1, X_2 and X_3.
- ....

This sampling procedure is quite good when the full distribution is to be sampled. If we want to sample from the conditional distribution of $X_2, \dots, X_n$ given X_1 then we can follow the same procedure, just fixing X_1 at the conditioned value. A similar procedure will work when conditioning on other X_i as it is quite easy to adjust the above procedure to start with any X_i.

When we want to sample from a conditional distribution with more than one conditioned variable the above strategy cannot be applied. An alternative is to use the

Gibbs sampler. This is possible since we have an explicit form for the density function. The only problem in computing the density is that some conditional distribution functions appear in the density function and these have to be recomputed. The highest degree vine seems to be the most convenient vine for this purpose since the number of conditional distribution functions that have to be computed is often small (largest when variable X_1 has been updated and smallest when variable X_n has been updated).

Recall that Gibbs sampling (for a thorough review in the context of Markov chain Monte Carlo methods see [6]) works by sampling from the conditional distributions in order. An arbitrary starting vector $(x_1, \dots, x_n)$ is chosen. Then we sample

- X_1 from the conditional distribution $F_{1|2\dots n}(\cdot; x_2, \dots, x_n)$. Call this sampled value $x_1^{(1)}$.
- X_2 from the conditional distribution $F_{2|13\dots n}(\cdot; x_1^{(1)}, x_3, \dots, x_n)$. Call this sampled value $x_2^{(1)}$.
- X_3 from the conditional distribution $F_{3|124\dots n}(\cdot; x_1^{(1)}, x_2^{(1)}, x_4, \dots, x_n)$. Call this sampled value $x_3^{(1)}$.
- ....
- X_n from the conditional distribution $F_{n|1\dots n-1}(\cdot; x_1^{(1)}, x_2^{(1)}, \dots, x_{n-1}^{(1)})$. Call this sampled value $x_n^{(1)}$.

This sampling procedure gives us a new vector $(x_1^{(1)}, \dots, x_n^{(1)})$. Clearly, the process for generating a new vector depends only on the previous vector. This means that we can view it as a discrete time Markov chain. Furthermore, since we have used the conditional distributions of $F_{1\dots n}$ it is clear that $F_{1\dots n}$ is a stationary distribution for the Markov process. Hence, if the Markov process is uniquely ergodic then $F_{1\dots n}$ is the only stationary distribution, and from the general theory of continuous state Markov chains (see, for example, [5]) we can conclude that the distribution of the state of the chain will converge to $F_{1\dots n}$ independently of the starting point.

Note that for conditional sampling we can apply the essentially the same procedure but with the conditioned variables held fixed.

In order to apply the Gibbs sampling method we need to show that the underlying Markov chain is uniquely ergodic. In [7] this was shown for the discrete case. In our situation, however, we are dealing with continuous random variables. For typical applications within uncertainty analysis we make use of marginal distributions elicited by expert judgement and rank correlations also given by experts. We are therefore particularly interested in this case.

From now on we assume that each variable X_i has compact support, that is, there is a compact interval I_i with full F_i probability and $f_i(x_i) > 0$ for all $x_i \in I_i$. A simple condition for the unique ergodicity of the Markov chain is that the density function is strictly positive on the product $I_1 \times \cdots \times I_n$.

Theorem 7. Let $\mathcal{V}$ be a rank-correlation vine specification. Suppose that the (conditional) rank correlations are uniformly bounded from -1 and $+1$, that is, there is a number α such that for all (conditional) rank correlations ρ we have

$$-1 < -\alpha \leqslant \rho \leqslant \alpha < 1.$$

Suppose also that all marginals F_i have support and positive density on compact intervals I_i. Then for the minimum information distribution with the given marginals and (conditional) rank correlation coefficients the Gibbs sampler is uniquely ergodic.

Proof. In [16] an expression is given for the density of the minimum information copula with given rank correlation. It is shown to take the form

$$\kappa(u,\theta)\kappa(v,\theta)\exp\left(\theta\left(u-\frac{1}{2}\right)\left(v-\frac{1}{2}\right)\right),$$

where $\theta = \theta(\rho)$ is a continuous monotone increasing function of the correlation coefficient. The function κ is strictly positive and is continuous as a function of θ (and thus also of τ). Therefore, since the rank correlations are uniformly bounded away from -1 and $+1$, the densities of the corresponding copulae are uniformly bounded from 0. Hence there are numbers $m > 0$ and $M > 0$ such that

$$m < f_i(x_i) < M \quad \text{for all } x_i \in I_i,$$

and

$$m < c(u,v) < M \quad \text{for all } (u,v) \in [0,1]^2 \text{ and all (conditional) copulae } c.$$

Using the density decomposition formula we then have

$$m^k < f_{1\ldots n} < M^k,$$

for some $k > 0$. For the conditional density $f_{1|2\ldots n} = f_{1\ldots n}/f_{2\ldots n}$ we get,

$$\frac{m^k}{M^k} < f_{1|2\ldots n} < \frac{M^k}{m^k}.$$

The same estimate holds for the other one-dimensional conditionals. This proves the theorem. □

We should remark that although the uniquely ergodic Markov chain always converges geometrically fast to the stationary distribution, it may still converge slowly in practice.

5.1. Conclusions

Regular vines give a simple way of specifying conditionally dependent random variables. The rank correlation vine specification is a particular case giving a highly convenient parameterization of a multivariate distribution. In particular, after specifying the marginal distributions, an expert specifies such a distribution by specifying $\binom{n}{2}$ numbers in $[-1, 1]$ which do not need satisfy any additional constraint.

We have shown that there is a density decomposition formula that generalizes the density decomposition for a Bayesian belief net using cliques. Using this density formula it is easy to derive the information decomposition theorem for regular vines.

Gibbs sampling appears to be a good way of sampling from unconditional and conditional vine distributions. In practical cases of rank correlation vines specified by expert assessment, simple conditions allowing the Gibbs sampler to be used will be in force.

Acknowledgement

We would like to thank Dorota Kurowicka for her helpful comments.

References

[1] T. Bedford and R.M. Cooke, Vines – a new graphical model for dependent random variables, Preprint, Delft University of Technology (1999), provisionally accepted for Annals of Statistics.
[2] R.M. Cooke, A.M.H. Meeuwissen and C. Preyssl, Modularizing fault tree uncertainty analysis: The treatment of dependent information sources, in: *Probabilistic Safety Assessment and Management*, ed. G. Apostolakis (Elsevier, 1991).
[3] R.M. Cooke, *UNICORN: Methods and Code for Uncertainty Analysis* (AEA Technologies, Warrington, 1995).
[4] R.M. Cooke, Markov and entropy properties of tree- and vine-dependent variables, in: *Proceedings of the ASA Section on Bayesian Statistical Science* (1997).
[5] J.L. Doob, *Stochastic Processes* (Wiley, 1953).
[6] D. Gamerman, *Markov Chain Monte Carlo; Stochastic Simulation for Bayesian Inference* (Chapman and Hall, London, 1997).
[7] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Pattern. Anal. Mach. Intell. 6 (1984) 721–741.
[8] R. Iman, J. Helton and J. Campbell, An approach to sensitivity analysis of computer models: Parts I and II, J. Quality Technol. 13(4) (1981).
[9] R. Iman and W. Conover, A distribution-free approach to inducing rank correlation among input variables, Comm. Statist. Simulation Comput. 11(3) (1982) 311–334.
[10] F.J. Jensen, *An Introduction to Bayesian Networks* (UCL Press, London, 1996).
[11] H. Joe, Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters, in: *Distributions with Fixed Marginals and Related Topics*, IMS Lecture Notes Monograph Series, Vol. 28, eds. L. Rüschendorf, B. Schweizer and M.D. Taylor (1996) pp. 120–141.
[12] S. Kullback, *Information Theory and Statistics* (Wiley, 1959).
[13] D. Kurowicka, Techniques in representing high dimensional distributions, Ph.D. Thesis, Delft University of Technology (2001).
[14] A.M.H. Meeuwissen, Dependent random variables in uncertainty analysis, Ph.D. Thesis, Delft University of Technology (1993).
[15] A.M.H. Meeuwissen and R.M. Cooke, Tree dependent random variables, Report 94-28, Department of Mathematics, Delft University of Technology (1994).

[16] A.M.H. Meeuwissen and T.J. Bedford, Minimal informative distributions with given rank correlation for use in uncertainty analysis, J. Statist. Comput. Simulation 57(1–4) (1997) 143–175.
[17] R.B. Nelsen, *An Introduction to Copulas*, Springer Lecture Notes in Statistics, Vol. 139 (Springer, New York, 1999).