Vines - a new graphical model for dependent random variables

Tim Bedford and Roger M. Cooke

Strathclyde University
Department of Management Science
40 George St, Glasgow G1 1QE, Scotland
tim@mansci.strath.ac.uk
and
Delft University of Technology
Faculty of Information Technology and Systems
Mekelweg 4, 2628 CD Delft, The Netherlands
r.m.cooke@its.tudelft.nl

21 August 2001

Keywords: correlation, dependence, information, multivariate probability distribution, Monte-Carlo simulation, tree dependence, Markov tree, belief net, multivariate normal distribution.
Mathematics Subject Classification: 62E10, 62E25, 62H20, 62B10, 94A17.

Abstract

A new graphical model, called a vine, for dependent random variables is introduced. Vines generalize the Markov trees often used in modelling high-dimensional 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.

Vines can be used to specify multivariate distributions in a straightforward way by specifying various marginal distributions and the ways in which these marginals are to be coupled. Such distributions have applications in uncertainty analysis where the objective is to determine the sensitivity of a model output with respect to the uncertainty in unknown parameters. Expert information is frequently elicited to determine some quantitative characteristics of the distribution such as (rank) correlations. We show that it is simple to construct a minimum information vine distribution, given such expert information. Sampling from minimum information distributions with given marginals and (conditional) rank correlations specified on a vine can be performed almost as fast as independent sampling. A special case of the

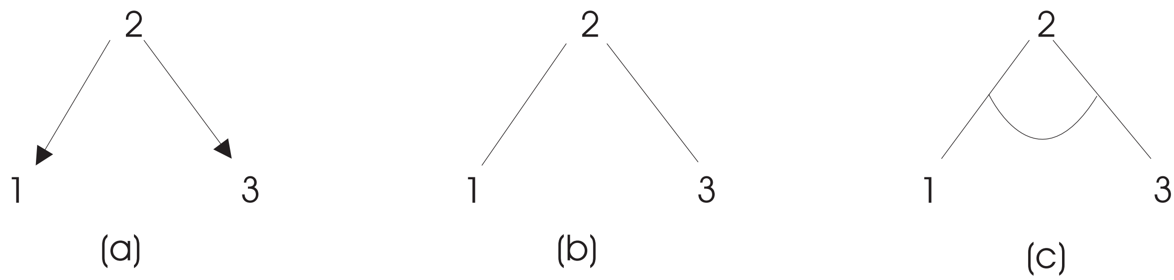

Fig. 1. *A belief net, a Markov tree, and a vine*

vine construction generalizes work of Joe and allows the construction of a multivariate normal distribution by specifying a set of partial correlations on which there are no restrictions except the obvious one that a correlation lies between -1 and 1.

1. Introduction Graphical dependency models have gained popularity in recent years following the generalization of the simple Markov trees to belief networks and influence diagrams. The main applications of these graphical models has been in problems of Bayesian inference with an emphasis on Bayesian learning (Markov trees and belief nets), and in decision problems (influence diagrams). Markov trees have also been used within the area of uncertainty analysis to build high-dimensional dependent distributions.

Within uncertainty analysis, the problem of easily specifying a coupling between two groups of random variables is prominent. Often, only some information about marginals is given (for example, some quantiles of a marginal distribution); extra information has to be obtained from experts, frequently in the form of correlation coefficients. In [3, 20, 21, 4], Markov trees are used to specify distributions used in uncertainty analysis (alternative approaches are found in [10, 11]). They 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. The use of the minimum information principle to motivate the use of a distribution with given correlation coefficient fits into the long-standing tradition established by Jaynes (see [12, 9]) in which subjective distributions are specified using moment information from an expert by maximizing entropy.

In this paper we show that the conditional independence property used in Markov trees and belief nets can be weakened without compromising ease of simulation. A new class of models called vines is introduced in which an expert can give input in terms of, for example, conditional rank correlations. 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. An important aspect is the ease with which the required information can be supplied by the expert - there are no joint restrictions on the correlations given (by contrast, for product moment correlations, the correlation matrix must always be positive definite). Our main result shows precisely how to obtain a minimum information vine distribution satisfying all the specifications of the expert.

Besides introducing the notion of a vine as a graphical model for conditional dependence, the paper shows how to construct joint distributions satisfying the conditional dependence specifications in a vine. The major element of this construction is the inductive generation of multivariate distributions with given marginals.

Sections 2 and 3 collect results for bivariate tree specifications. Section 4 introduces a more general type of specification in which conditional marginal distributions can be stipulated or qualified. The tree structure for bivariate constraints generalizes to a "vine" structure for conditional bivariate constraints. A vine is a sequence of trees such that the edges of tree T_i are the nodes of T_{i+1}. Minimum information results show that complicated conditional independence properties can be obtained from vine specifications in combination with information minimization. Sampling from minimum information distributions given marginal and (conditional) rank correlations specified on a vine can be performed at a speed comparable to independent sampling.

A vine is a convenient tool with a graphical representation that makes it easy to describe which conditional specifications are being made for the joint distribution. The existence of distributions satisfying these constraints is proven more easily by generalizing the construction to Cantor trees, as is done in Section 5. The existence of joint distributions satisfying Cantor tree specifications is shown, and a formula for the information of such a distribution (relative to the independent distribution with the same marginals) is proven. This section also contains a particular way of constructing joint distributions from given, overlapping, marginals. Section 6 shows that the regular vines are special cases of Cantor tree constructions, and that Cantor trees can be represented graphically by vines. Finally, Section 7 gives specific results for rank and partial correlation specifications. It is shown that for these hierarchical constructions there are no restrictions on rank or partial correlation specifications, except for the obvious one that correlation must be between -1 and 1. In particular, a joint normal distribution can be specified without worrying about positive definiteness considerations.

Sections 2 to 4 are based on, or developed directly from [5].

The general topic addressed in this paper, that of specifying a distribution with given marginals, has been addressed elsewhere. In particular, Li et al [18, 19] develop alternative ways of coupling distibutions on overlapping sets of variables. Joe [13, 14] gives a number of methods for generating distributions with given marginals. In particular the construction of Section 4.5 in [14] corresponds to the most simple type of vine as shown in Figure 2. In the appendix to [13] he uses this same type of simple vine structure to specify a multivariate normal distribution - a construction that we call the standard vine, and that we generalize. Other authors have looked at alternative ways of specifying multivariate distributions. For example [1] gives a survey of methods in which conditional distributions are used to define, or at least partially specify, the multivariate distribution.

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 $(1 \leq i \leq n)$, the bivariate distribution functions are F_{ij} $(1 \leq i \neq j \leq n)$, and $F_{i|j}$ denotes the distribution of variable i conditional on j. The same subscript conventions apply to densities f and laws μ. Whenever we use the relative information integral, the absolute continuity condition mentioned below is assumed to hold.

Definition 1 relative information.
Let ν and μ be probability measures on a probability space such that ν is absolutely continuous with respect to μ with Radon-Nikodym derivative $\frac{d\nu}{d\mu}$, then the relative information *or* Kullback-Liebler divergence, *$I(\nu|\mu)$ of ν with respect to μ is*

$$I(\nu|\mu) = \int \log(\frac{d\nu}{d\mu}(x))\ d\nu(x)\ .$$

When ν is not absolutely continuous with respect to μ we define $I(\nu|\mu) = \infty$.

In this paper we shall construct distributions that are as "independent" as possible given the 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 $I(\nu|\mu)$ can be interpreted as measuring the degree of "uniformness" of ν (with respect to μ). The relative information is always non-negative and equals zero if and only if $\mu = \nu$. See for example [17] and [8].

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*

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

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

$$\rho(U, V) = \mathrm{cov}(U, V)/\sqrt{\mathrm{var}(U)\mathrm{var}(V)},$$

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

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

where $(\tilde{X}_1, \tilde{X}_2)$ 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 measure 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 [22]. 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.

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).*

Note that we do not assume that T is connected. 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. *$B = \{B(i, j)|\{i, j\} \in E\}$, 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 $(1 \leq i \leq n)$ 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 whenever $m \geq M$ and $i, j \in N$ are joined by edges $\{i, k_1\}, \ldots, \{k_m, j\} \in E$ we have that X_i and X_j are conditionally independent given any M of k_ℓ, $1 \leq \ell \leq m$; 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 $M > 0$.*

One approach, implemented for example in [16], is 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 $t = \{t_{ij} | t_{ij} \in [-1,1],\ \{i,j\} \in E,\ t_{ij} = t_{ji},\ t_{ii} = 1\}$.*

The following three results are proved in [21]. The first is similar to results about influence diagrams [24], the second uses a construction of [6].

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

$$g(x_1, \ldots, x_n) = \frac{\prod_{(i,j) \in E} f_{ij}(x_i, x_j)}{\prod_{i \in N} (f_i(x_i))^{d(i)-1}}\ , \tag{2.1}$$

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

The following theorem states that a rank correlation tree specification is always consistent.

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.

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.

The multivariate probability distribution function $F_{\underline{X}}$ of any random vector $\underline{X}$ can be obtained as the n-dimensional marginal distribution of a realization of a bivariate tree specification of an enlarged vector $(\underline{X}, \mathcal{L})$.

Theorem 3. *Given a vector of random variables $\underline{X} = (X_1, \ldots, X_n)$ with joint probability distribution $F_{\underline{X}}(\underline{x})$, there exists an $(n+1)$-dimensional bivariate tree specification (G, T, B) on random variables $(Z_1, \ldots, Z_n,\ \mathcal{L})$ whose distribution $G_{\underline{Z},\mathcal{L}}$ is Markov tree dependent, such that $\int G_{\underline{Z},\mathcal{L}}(\underline{x}, \ell)\ d\ell = F_{\underline{X}}(\underline{x})$.*

3. 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,

$$I(g|\prod_{i\in N} f_i) = \sum_{\{i,j\}\in E} I(f_{ij}|f_i f_j)\ .$$

If the bivariate tree specification does not completely specify the bivariate marginals $f_{ij}, \{i,j\} \in E$, then more than one Markov tree dependent realization may be possible. In this case 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 $I(f_{ij}|f_i f_j), \{i,j\} \in E$.

In this section we show that 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, we show that a minimal information realization of a bivariate tree specification has Markov tree dependence. This follows from a very general result (Theorem 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.

To prove this theorem, we first formulate three lemmas. We assume in this analysis that the distributions have densities. Throughout this section, Z, Y and X are finite dimensional random vectors having no components in common. To recall notation, $g_{X,Y,Z}(x,y,z)$ is a density with marginal densities $g_X(x)$, $g_Y(y)$, $g_Z(z)$, and bivariate marginals $g_{X,Y}$, $g_{X,Z}$ and $g_{Y,Z}$. We write $g_{X|Y}$ for the conditional density of X given Y.

Lemma 1. *Let $g_{X,Y,Z}$ be a density and define*

$$\tilde{g}_{X,Y,Z}(x,y,z) = g_X(x) g_{Y|X}(x,y) g_{Z|X}(x,z)$$

Then $\tilde{g}_{X,Y,Z}$ satisfies

$$\tilde{g}_X = g_X\ ,\quad \tilde{g}_Y = g_Y\ ,\quad \tilde{g}_Z = g_Z\ ,$$

$$\tilde{g}_{X,Y} = g_{X,Y}\ ,\quad \tilde{g}_{X,Z} = g_{X,Z}$$

and makes Y and Z conditionally independent given X.

Proof: The proof is a straightforward calculation. □

Lemma 2. *With g as above, let $p_X(x)$ be a density. Then*

$$\int g_Y(y) I(g_{X|Y}|p_X)\ dy \geq I(g_X|p_X),$$

with equality holding if and only if X and Y are independent under g; that is, if $g_{X|Y}(x,y) = g_X(x)$.

Proof: By definition,

$$\int g_Y(y) I(g_{X|Y}|p_X)\ dy \geq I(g_X|p_X)$$

is equivalent to

$$\int\int g_Y(y) g_{X|Y}(x,y) \log \frac{g_{X|Y}(x,y)}{p_X(x)}\ dxdy \geq \int g_X(x) \log \frac{g_X(x)}{p_X(x)}\ dx$$

and hence to

$$\int\int g_{X,Y}(x,y) \log g_{X|Y}(x,y)\ dxdy \geq \int\int g_{X,Y}(x,y) \log g_X(x)\ dxdy\ .$$

This can be rewritten as

$$\int\int g_{X,Y}(x,y) \log \frac{g_{X|Y}(x,y)}{g_X(x)}\ dxdy \geq 0$$

or equivalently

$$\int\int g_{X,Y}(x,y) \log \frac{g_{X,Y}(x,y)}{g_X(x) g_Y(y)}\ dxdy \geq 0\ . \tag{3.2}$$

The left side of the last inequality equals $I(g_{X,Y}|g_X g_Y)$. Inequality 3.2 always holds and it holds with equality if and only if $g_{X,Y} = g_X g_Y$ (see [17]). □

Remark: The quantity on the left side of Equation 3.2 is also called *mutual information.*

Lemma 3. *Let $g_{X,Y,Z}(x,y,z)$ and $\tilde{g}_{X,Y,Z}(x,y,z)$ be two probability densities defined as in Lemma 1. Then*

i) $I(g_{X,Y,Z}|g_X g_Y g_Z) \geq I(\tilde{g}_{X,Y,Z}|g_X g_Y g_Z)$,

ii) $I(\tilde{g}_{X,Y,Z}|g_X g_Y g_Z) = I(g_{X,Y}|g_X g_Y) + I(g_{X,Z}|g_X g_Z)$.

Equality holds in (i) if and only if $g = \tilde{g}$.

Proof: By definition we have

$$I(g_{X,Y,Z}|g_X g_Y g_Z) = \int\int\int g_{X,Y,Z}(x,y,z) \log \frac{g_{X,Y,Z}(x,y,z)}{g_X(x) g_Y(y) g_Z(z)}\ dxdydz$$

which by conditionalization is equivalent with

$$\int\int\int g_{X,Y,Z}(x,y,z) \log \frac{g_{X,Y}(x,y) g_{Z|X,Y}(x,y,z)}{g_X(x) g_Y(y) g_Z(z)}\, dxdydz =$$

$$= I(g_{XY}|g_X, g_Y) + \int\int\int g_{X,Y,Z}(x,y,z) \log \frac{g_{Z|X,Y}(x,y,z)}{g_Z(z)}\, dxdydz.$$

The second term can be written as

$$\int\int\int g_{X,Y}(x,y) g_{Z|XY}(z) \log \frac{g_{Z|X,Y}(x,y,z)}{g_Z(z)} dzdxdy =$$

$$= \int\int g_{X,Y}(x,y) I(g_{Z|XY}|g_Z) dxdy =$$

$$= \int g_X \int g_{Y|X}(x,y) I(g_{Z|XY}|g_Z) dydx \geq \int g_X I(g_{Z|X}|g_Z)\ dx =$$

$$= \int\int g_X g_{Z|X} \log \frac{g_{Z|X}(z) g_X(x)}{g_Z(z)\, g_X(x)}\ dzdx = I(g_{XZ}|g_X\ g_Z)$$

where Lemma 2 is used for the inequality. Hence

$$I(g_{X,Y,Z}|g_X g_Y g_Z) \geq I(g_{XY}|g_X g_Y) + I(g_{XZ}|g_X g_Z) \tag{3.3}$$

with equality if and only if Z and Y are independent given X, which holds for $\tilde{g}$ (Lemma 1). □

We may now formulate

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

Proof: Let $f_{X,Y,Z}$ be a joint probability density with marginals f_X, f_Y, f_Z, whose two dimensional marginals satisfy the constraints $B(X,Y)$ and $B(X,Z)$. Assume that f satisfies $I(f_{X,Y,Z}|f_X f_Y f_Z) \leq I(g_{X,Y,Z}|f_X f_Y f_Z)$. Then by Lemma 1 and Lemma 3(i) we may assume without loss of generality that $f_{X,Y,Z} = \tilde{f}_{X,Y,Z} := f_{XY} f_{Z|X}$. By Lemma 3(ii) we have

$$I(\tilde{f}_{X,Y,Z}|f_X f_Y f_Z) = I(f_{X,Y}|f_X f_Y) + I(f_{X,Z}|f_X f_Z).$$

But

$$\begin{aligned} I(f_{X,Y}|f_X f_Y) + I(f_{X,Z}|f_X f_Z) &\geq I(g_{X,Y}|f_X f_Y) + I(g_{X,Z}|f_X f_Z) = \\ &= I(g_{X,Y,Z}|f_X f_Y f_Z) \geq \\ &\geq I(\tilde{f}_{X,Y,Z}|f_X f_Y f_Z) = \\ &= I(f_{X,Y}|f_X f_Y) + I(f_{X,Z}|f_X f_Z). \end{aligned}$$

By the uniqueness of $g_{X,Z}$ and $g_{X,Y}$, this entails $g_{X,Y,Z} = f_{X,Y,Z}$. □

Corollary 1. *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 $\{i,j\} \in E$.*

Proof: Using the notation of Theorem 1, the proof is by induction on n. For $n = 2$ there is nothing to prove. For $n = 3$ the result follows from Lemma 3(ii).
Assume now that we have a tree with $n+1$ nodes. Assume also that there is a node with degree 1 (otherwise all nodes have degree 0, there are no constraints and the result holds trivially). Let Z be the variable corresponding to this node, X the variable corresponding to its unique neighbour, and Y the vector of variables corresponding to the other $n-1$ nodes. Applying the Lemma 3(i)

we see that the information is minimized by the distribution making Y and Z conditionally independent given X. Since by induction the marginal g_{XY} is minimally informative, Lemma 3(ii) implies that g_{XZ} also must be minimally informative as claimed. □

If $B(i,j)$ fully specifies $g(x_i, x_j)$ for $\{i,j\} \in E$, then the above corollary says that there is a unique minimum information density given $(\underline{F}, T, B)$ and this density is Markov tree dependent.

4. Regular vines

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. We seek a more general structure for partially specifying joint distributions and obtaining minimal information results. 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:

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

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 [1]. 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 2 and 3 show regular vine specifications on 5 variables. Figure 2 corresponds to the structure studied by Joe [13]. Each edge of a regular vine is associated with a restriction on the bivariate or conditional bivariate distribution shown adjacent to the edge.
Note that the bottom level restrictions on the bivariate marginals form a tree T_1 with nodes $1, \dots, 5$. 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, $i = 1, \dots, n-1$ have maximal cardinality $n - i$, as we may "remove" any edge by associating with it the vacuous restriction.
A regular vine is a special case of 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. In this section we define the notion of a regular vine. The more general concept of a vine will be developed in the next section, together with existence and uniqueness results for distributions satisfying vine constraints.

DEFINITION 8 REGULAR VINE, VINE. *$\mathcal{V}$ is a* vine *on n elements if*

1. $\mathcal{V} = (T_1, ..., T_m)$
2. *T_1 is a tree with nodes $N_1 = \{1, \dots, n\}$ and a set of edges denoted E_1,*

[1] We ignore measurability constraints here, but return to discuss them later.

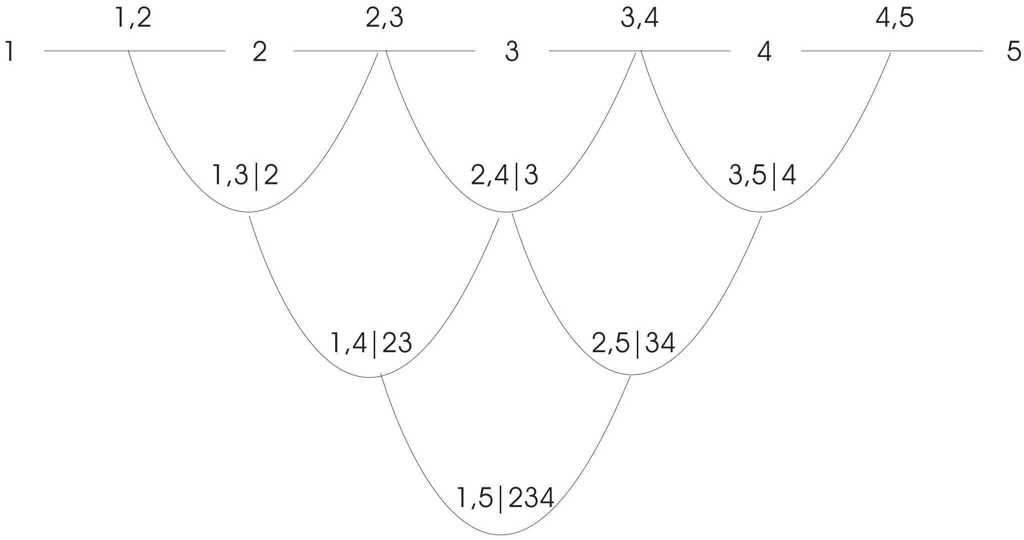

FIG. 2. *A regular vine*

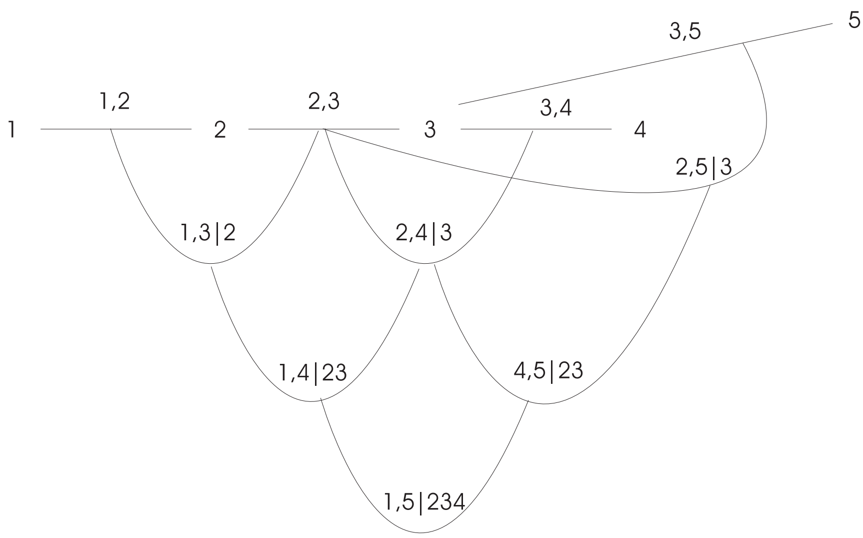

FIG. 3. *Another regular vine*

3. *For* $i = 2, \ldots, m$, T_i *is a tree with nodes* $N_i \subset N_1 \cup E_1 \cup E_2 \ldots \cup E_{i-1}$ *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* $i = 1, \ldots, n$, *where* $\#N_i$ *is the cardinality of the set* N_i.
3. *The proximity condition holds: For* $i = 2, \ldots, n-1$, *if* $a = \{a_1, a_2\}$ *and* $b = \{b_1, b_2\}$ *are two nodes in* N_i *connected by an edge (recall* $a_1, a_2, b_1, b_2 \in N_{i-1}$*), then* $\#a \cap b = 1$.

It will be convenient to introduce some labeling corresponding to the edges and nodes in a vine, in order to specify the constraints. In order to do this 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 $N_i \subset E_0 \cup E_1 \cup E_2 \ldots \cup E_{i-1}$ (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*

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

For a regular vine and an edge $e_i \in E_i$ *the* j-fold union *of* e_i $(0 < j \leq i-1)$ *is the subset*

$$U_{e_i}(j) = \{e_{i-j} \in E_{i-j} | \exists \text{ edges } e_k \in E_k, (k = i-j+1, \ldots, i-1), \\ \text{with } e_k \in e_{k+1} (k = i-j, \ldots, i-1)\}.$$

For $j = 0$ *define* $U_{e_i}(0) = \{e_i\}$.

We can now define the constraint sets.

Definition 10 constraint set. *For* $e = \{j, k\} \in E_i$, $i = 1, \ldots, m-1$, *the* conditioning set *associated with* e *is*

$$D_e = A_j \cap A_k,$$

and the conditioned sets *associated with* e *are*

$$C_{e,j} = A_j - D_e, \quad \text{and } C_{e,k} = A_k - D_e.$$

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

$$\mathcal{CV} = \{(C_{e,j}, C_{e,k}, D_e) | i = 1, \ldots, m-1, e \in E_i, e = \{j, k\}\}.$$

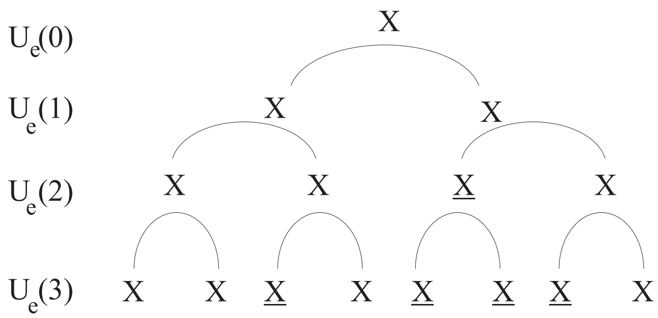

Fig. 4. *Counting edges*

Note that $A_e = A_j \cup A_k = C_{e,j} \cup C_{e,k} \cup D_e$ when $e = \{j,k\}$. For $e \in E_m$ the conditioning set is empty.
The constraint set is shown for the regular vines in Figures 2 and 3. At each edge $e \in E_i$, the terms $C_{e,j}$ and $C_{e,k}$ are separated by a comma and given to the left of the "|" sign, while D_e appears on the right. For example, in Figure 2, 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 the rest of this section we shall discuss properties of regular vines. The existence of distributions corresponding to regular vines will be dealt with in a later section on vines.

Lemma 4. *Let $\mathcal{V}$ be a regular vine on n elements, and let $e \in E_i$. Then $\#U_e(j) = j+1$ for $j = 0,1,\ldots,i$.*

Proof: The statement clearly holds for $j = 0$ and $j = 1$. By the proximity property it follows immediately that it holds for $j = 2$. We claim that in general

$$\#U_e(j) = 2\#U_e(j-1) - \#U_e(j-2), \quad j = 2,3,\ldots,$$

after which the result follows by induction. To see this we represent the $U_e(j)$ as a complete binary tree whose nodes are in a set of nodes of $\mathcal{V}$. The repeated nodes are underscored, and children of underscored nodes are underscored. Because of proximity, nodes with a common parent must have a common child. Letting X denote an arbitrary node we have the situation shown in Figure 4.
Evidently the number of newly underscored nodes on echelon k (that is, nodes which are not children of an underscored node) is equal to the number of non-underscored nodes in echelon $k-2$. Hence, the number of non-underscored nodes in echelon k is $2\#U_e(k-1) - \#U_e(k-2)$. $\square$

Lemma 5. *If $\mathcal{V}$ is a regular vine on n elements then for all $i = 1,\ldots,n-1$, and all $e \in E_i$ the conditioned sets associated with e are singletons, $\#C_{e,j} = 1$. Furthermore, $\#A_e = i+1$, and $\#D_e = i-1$.*

Proof: By Lemma 4 we have $\#A_e = i+1$. The proof of the other claims is by induction on $i = 1, \ldots, n-1$. The statements clearly hold for $i = 1$. Suppose they hold for m, $1 \leq m < i$. Let $e = \{j, k\}$, where $j = \{j_1, j_2\}$ and $k = \{k_1, k_2\}$. By the proximity property one of j_1, j_2 equals one of k_1, k_2, say $j_1 = k_1$. We have

$$A_e = A_{j_1} \cup A_{j_2} \cup A_{k_1} \cup A_{k_2}.$$

By induction,

$$\#D_j = \#(A_{j_1} \cap A_{j_2}) = i-2,$$

and $\#A_{j_1} = \#A_{j_2} = i-1$ and

$$\#A_j = \#(A_{j_1} \cup A_{j_2}) = i.$$

Hence $A_{j_2} - A_{j_1}$ contains exactly one element, and similarly for $A_{k_2} - A_{k_1}$. Moreover, these two elements must be distinct, since otherwise $A_j = A_k$, which would imply that $\#A_e = i$ in contradiction of Lemma 4. Hence

$$\#A_e = \#(A_j \cup A_k) = i+1, \quad \#D_e = i-1, \quad \text{and} \quad D_e = A_{j_1} = A_{k_1}.$$

□

LEMMA 6. *Let $\mathcal{V}$ be a regular vine on n elements and $j, k \in E_i$. Then $A_j = A_k$ implies $j = k$.*

Proof: Suppose not. Then there is a largest x such that $U_j(x) \neq U_k(x)$ and $U_j(x+1) = U_k(x+1)$. Since $\#U_j(x+1) = x+2$ there can be at most $x+1$ edges between the elements of $U_j(x+1)$ in the tree T_{i-x-1}. But since $\#U_j(x) = \#U_k(x) = x+1$ we must have that $U_j(x) = U_k(x)$ because otherwise this would contradict T_{i-x-1} being a tree. □

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 invertable distribution functions.*
2. *$\mathcal{V}$ is a regular vine on n elements*
3. *$B = \{B_e(d) | i = 1, \ldots n-1;\, e \in E_i\}$ 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* realise 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 $(i = 1, \ldots, n)$.*

We shall see later that regular vine dependent distributions can be constructed. However, in order to construct distributions (as opposed to simply constrain distributions as we do in the above definition) it is necessary to make an additional measurability assumption. This is that for any edge e, for any Borel set $B \subset [0,1]^2$, the copula measure of B given X_{D_e} is a measurable function of X_{D_e}. A family of conditional copulae indexed by X_{D_e} with this property is called a *regular conditional copulae family.*

A convenient way, but not the only 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 will be discussed in Section 7.

The existence of regular vine distributions will follow from more general result given in the next section, but we illustrate briefly how such a distribution is determined using the regular vine in Figure 2 as an example. We make use of the expression

$$g_{12345} = g_1 g_{2|1} g_{3|12} g_{4|123} g_{5|1234}.$$

The marginal distribution of X_1 is known, so we have g_1. The marginals of X_1 and X_2 are known, and the copula of X_1, X_2 is also known, so we can get g_{12}, and hence $g_{2|1}$. In order to get the third term $g_{3|12}$ we determine $g_{3|2}$ similarly to $g_{2|1}$. Next we calculate $g_{1|2}$ from g_{12}. With $g_{1|2}$, $g_{3|2}$, and the conditional copula of X_1, X_3 given X_2 we can determine the conditional joint distribution $g_{13|2}$, and hence the conditional marginal $g_{3|12}$. Progressing in this way we obtain $g_{4|123}$ and $g_{5|1234}$.

We note that a regular vine on n elements is uniquely determined if the nodes N_1 have degree at most 2 in T_1. If T_1 has nodes of degree greater than 2, then there is more than one regular vine. Figure 2 shows a regular vine that is uniquely determined, the regular vine in Figure 3 is not uniquely determined. The edge labelled [25|3] could be replaced by an edge [45|3].

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. The results needed to do this are contained in the following lemma.

Recalling our standard notation, and moving from densities to general Borel probability measures, μ is a Borel probability measure on $\mathbb{R}^n$, $\mu_{1,..k}$ denotes the marginal over $x_1, ...x_k$, $\mu_{1,..k-1|k,..n}$ denotes the marginal over $x_1, ...x_{k-1}$ conditional on $x_k, ...x_n$. Finally, $E_{1,...k}$ denotes expectation taken over $x_1, ...x_k$ taken with respect to $\mu_{1,..,k}$.

The following lemma contains useful facts for computing with relative information for multivariate distributions. The proof is similar in spirit to the proofs of the previous section, and will be indicated summarily here.

Lemma 7. *Suppose that $I(\mu|\prod_{i=1}^n \mu_i) < \infty$, then:*

1.

$$I(\mu|\prod_{i=1}^{n} \mu_i) = I(\mu_{k,...n}|\prod_{i=k}^{n} \mu_i) + E_{k,...n} I(\mu_{1,...k-1|k,...n}|\prod_{i=1}^{k-1} \mu_i).$$

2.

$$I(\mu|\prod_{i=1}^{n}\mu_i) = \sum_{j=1}^{n-1} E_{1,\dots j} I(\mu_{j+1|1,\dots j}|\mu_{j+1}).$$

3.

$$\begin{aligned} &E_{2,\dots n} I(\mu_{1|2\dots n}|\mu_1) + E_{1\dots n-1} I(\mu_{n|1\dots n-1}|\mu_n) = \\ &\quad = E_{2,\dots n-1}\left(I(\mu_{1,n|2,\dots n-1}|\mu_{1|2,\dots n-1}\mu_{n|2,\dots n-1}) + I(\mu_{1,n|2,\dots n-1}|\mu_1\mu_n)\right) \end{aligned}$$

4.

$$\begin{aligned} &2I(\mu|\prod_{i=1}^{n}\mu_i) = I(\mu_{2,\dots n}|\prod_{i=2}^{n}\mu_i) + I(\mu_{1,\dots n-1}|\prod_{i=1}^{n-1}\mu_i) + \\ &\quad + E_{2,\dots n-1} I(\mu_{1,n|2,\dots n-1}|\mu_{1|2,\dots n-1}\mu_{n|2,\dots n-1}) + I(\mu|\mu_1\mu_n\mu_{2,\dots n-1}) \end{aligned}$$

5.

$$\begin{aligned} &I(\mu|\prod_{i=1}^{n}\mu_i) = I(\mu_{2,\dots n}|\prod_{i=2}^{n}\mu_i) + I(\mu_{1,\dots n-1}|\prod_{i=1}^{n-1}\mu_i) \\ &\quad - I(\mu_{2,\dots n-1}|\prod_{i=2}^{n-1}\mu_i) + E_{2,\dots n-1} I(\mu_{1,n|2,\dots n-1}|\mu_{1|2,\dots n-1}\mu_{n|2,\dots n-1}). \end{aligned}$$

Proof: We indicate the main steps, leaving the computational details to the reader. Since $I(\mu|\prod_{i=1}^{n}\mu_i) < \infty$ there is a density g for μ with respect to $\prod_{i=1}^{n}\mu_i$. We use the usual notation for the marginals etc of g.

1. For μ on the left hand side fill in $g = g_{1,\dots k-1|k,\dots n} g_{k,\dots n}$.
2. This follows from the above by iteration.
3. The integrals on the left hand side can be combined, and the logarithm under the integral has the argument:

   $$\frac{gg}{g_{2,\dots n} g_{1,\dots n-1} g_1 g_n}.$$

   This can be re-written as

   $$\frac{g_{1,n|2,\dots n-1}}{g_{1|2,\dots n-1} g_{n|2,\dots n-1}} \frac{g_{1,n|2,\dots n-1}}{g_1 g_n}.$$

   Writing the log of this product as the sum of logarithms of its terms, the result on the right hand side is obtained.
4. This follows from the first and the previous statement by noting

   $$E_{2,\dots n-1} I(\mu_{1,n|2,\dots n-1}|\mu_1\mu_n) = I(\mu|\mu_1\mu_n\mu_{2,\dots n-1}).$$

5. This follows from the previous two statements by noting

   $$I(\mu|\prod_{i=1}^{n}\mu_i) = I(\mu|\mu_1\mu_n\mu_{2,\dots n-1}) + I(\mu_{2,\dots n-1}|\prod_{i=2}^{n-1}\mu_i).$$

□

As an example consider the regular vine shown in Figure 2. We have,

$$
\begin{aligned}
I(\mu_{12345}|\mu_1\dots\mu_5) &= I(\mu_{1\dots4}|\prod_{i=1}^{4}\mu_i) + I(\mu_{2\dots5}|\prod_{i=2}^{5}\mu_i) \\
&\quad -I(\mu_{2\dots4}|\prod_{i=2}^{4}\mu_i) + E_{2\dots4}I(\mu_{1,5|2\dots4}|\mu_{1|2\dots4}\mu_{5|2\dots4}) \\
&= I(\mu_{123}|\prod_{i=1}^{3}\mu_i) + I(\mu_{234}|\prod_{i=2}^{4}\mu_i) + I(\mu_{345}|\prod_{i=3}^{5}\mu_i) + \\
&\quad -I(\mu_{23}|\mu_2\mu_3) - I(\mu_{34}|\mu_3\mu_4) + \\
&\quad E_{23}I(\mu_{1,4|23}|\mu_{1|23}\mu_{5|23}) + E_{34}I(\mu_{2,5|34}|\mu_{2|34}\mu_{5|34}) + \\
&\quad E_{234}I(\mu_{1,5|234}|\mu_{1|234}\mu_{5|234}) \\
&= I(\mu_{12}|\mu_1\mu_2) + I(\mu_{23}|\mu_2\mu_3) + I(\mu_{34}|\mu_3\mu_4) + I(\mu_{45}|\mu_4\mu_5) + \\
&\quad E_2I(\mu_{13|2}|\mu_{1|2}\mu_{3|2}) + E_3I(\mu_{24|3}|\mu_{2|3}\mu_{4|3}) + E_4I(\mu_{35|4}|\mu_{3|4}\mu_{5|4}) + \\
&\quad E_{23}I(\mu_{1,4|23}|\mu_{1|23}\mu_{5|23}) + E_{34}I(\mu_{2,5|34}|\mu_{2|34}\mu_{5|34}) + \\
&\quad E_{234}I(\mu_{1,5|234}|\mu_{1|234}\mu_{5|234})
\end{aligned}
$$

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. The calculation can be generalized to all regular vines, as is shown in the next result. As it is a special case of a more general result, the Information Decomposition Theorem, to be given in the next section, we give no proof.

THEOREM 5. *Let μ be a Borel probability measure on $\mathbb{R}^n$ satisfying the regular vine specification $(\underline{F}, \mathcal{V}, B)$, and suppose that for each i, $e = \{j,k\} \in E_i$, and $d \in D_e$, $\mu_{C_{e,j},C_{e,k}|d}$ is a Borel probability measure minimizing*

$$I(\mu_{C_{e,j},C_{e,k}|d}|\mu_{C_{e,j}|d}\mu_{C_{e,k}|d}). \tag{4.4}$$

Then μ satisfies $(\underline{F}, \mathcal{V}, B)$ and minimizes

$$I(\mu|\prod_{i=1}^{n}\mu_i). \tag{4.5}$$

Furthermore, if any of the $\mu_{C_{e,j},C_{e,k}|d}$ uniquely *minimizes the information term in Expression 4.4 (for all values d of D_e), then μ minimizes the information term in Expression 4.5.*

5. Cantor specifications and the Information Decomposition Theorem

The definition of a regular vine can be generalized to that of a vine to allow a wider variety of constraints than is possible with a regular vine. The main problem we then face, however, is that arbitrary specifications might not be consistent. The situation is analogous to that for a product-moment correlation matrix where the entries can be taken arbitrarily between -1 and 1 but have to satisfy the additional (global) constraint of positive

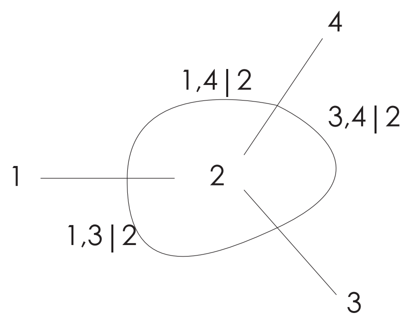

FIG. 5. *Inconsistent vine representation*

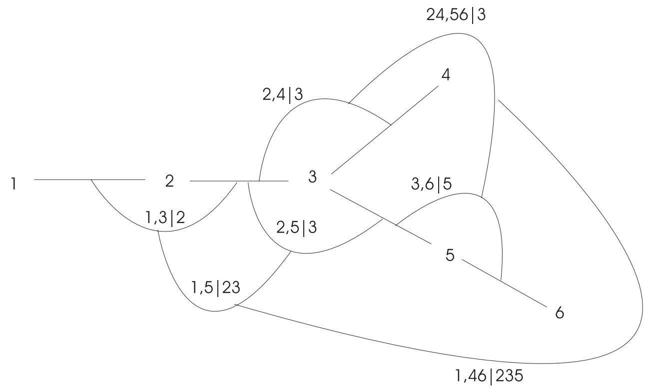

FIG. 6. *Another inconsistent vine representation*

definiteness. We wish to define a graphical structure so that one can build a multivariate distribution by specifying functionally independent properties encoded by each node on a vine. Furthermore, we wish to define a general structure that allows the decomposition of the information in a similar way to that given in Theorem 5.

An example of the problems that can arise when one attempts to generalize the definition of a regular vine is shown in Figure 5. This figure shows a vine with a cycle of constraints giving, for example, two specifications of the distribution of (X_1, X_2, X_4) which need not be consistent. This example is a vine under the definition given in the last section: Take T_1 with edge set $\{e_1 = \{1,2\}, e_2 = \{2,4\}, e_3 = \{2,3\}\}$, T_2 with edge set $\{\{e_1, e_3\}, \{e_1, e_2\}\}$, and T_3 with edge set $\{\{e_2, e_3\}\}$. An example that allows an inconsistent specification but that contains no cycles is given in Figure 6. Here, the joint distribution of (X_2, X_3, X_5) is specified in two distinct ways, by the $2,5|3$ and the $24,56|3$ branch.

We shortly give another approach to building joint distributions that will avoid this problem, and which allow us to build vines sustaining distributions. This second approach is a "top-down" construction called a Cantor tree (as compared with the "bottom-up" vine construction).

We first give a general definition of a coupling that enables us to define joint distributions for pairs of random vectors. Recall that the usual definition of a copula is as a distribution on the unit square with uniform marginals. A copula is used to couple two random variables in such a way that the marginals are preserved. Precisely, if X_1 and X_2 are random variables with distribution functions F_{X_1} and F_{X_2}, and if C is the distribution function of a copula then

$$(x_1, x_2) \mapsto C(F_{X_1}(x_1), F_{X_2}(x_2)) \tag{5.6}$$

is a joint distribution function with marginals F_{X_1} and F_{X_2}.

Definition 13. *Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two measurable spaces, and $\mathcal{P}(S, \mathcal{S})$ and $\mathcal{P}(T, \mathcal{T})$ the sets of probabilities on these spaces. A* coupling *is a function*

$$C : \mathcal{P}(S, \mathcal{S}) \times \mathcal{P}(T, \mathcal{T}) \to \mathcal{P}(S \times T, \mathcal{S} \otimes \mathcal{T})$$

(where $\mathcal{S} \otimes \mathcal{T}$ denotes the product sigma-algebra), with the property that for any $\mu \in \mathcal{P}(S, \mathcal{S})$, $\nu \in \mathcal{P}(T, \mathcal{T})$ the marginals of $C(\mu, \nu)$ are μ and ν respectively.

Genest *et al* [7] show that the natural generalization of Equation 5.6, in which the X_i are replaced by vectors X_i (and F_{X_i} by multivariate distribution functions), cannot work unless C is the independent copula because the function defined in this way is not in general a multivariate distribution function. Hence we have to find a different way of generalizing the usual construction of a copula. Here we give one approach. There are other approaches, for example discussed in [18] and [19]. We assume that all spaces are Polish, to be able to decompose measures into conditional measures.

Definition 14. *Let μ_1 and μ_2 be probability distributions supported on probability spaces V_1 and V_2, and let $\varphi_i : V_i \to \mathbb{R}$ $(i = 1, 2)$ be Borel measurable functions. If c is a copula then the* $(\varphi_1, \varphi_2, c)$-coupling *for μ_1 and μ_2 is the probability distribution μ on $V_1 \times V_2$ defined as follows: Let F_i be the distribution function of the probability $\mu_i \circ \varphi_i^{-1}$, and denote by $\mu_{i|u}$ the conditional probability distribution of μ_i given $F_i \varphi_i = u$. Then μ is the unique probability measure such that*

$$\int f(v_1, v_2)\, d\mu(v_1, v_2) = \int \int \int f(v_1, v_2)\, d\mu_{1|u_1}(v_1) d\mu_{2|u_2}(v_2) dc(u_1, u_2),$$

for any characteristic function f of a measurable set $B \subset V_1 \times V_2$.

Remark: An alternative way to construct a random vector (X_1, X_2) with distribution μ is as follows: Define (U_1, U_2) to be random variables in the unit square with distribution c. Let F_i be the distribution function of a random variable $\varphi_i(Y_i)$ where Y_i has distribution μ_i. Then, given $U_i = u_i$, define X_i to be independent of X_{3-i} and U_{3-i} with the distribution of Y_i conditional on $F_i \varphi_i(Y_i) = u_i$ $(i = 1, 2)$. This is shown in the Markov tree in Figure 7.

It is easy to see that the marginals of the $(\varphi_1, \varphi_2, c)$-coupling are μ_1 and μ_2. We have therefore defined a coupling in the sense of Definition 13.
Clearly we could take φ_i to be the distribution function of μ_i when V_i is a subset of Euclidean space. When additionally $V_1, V_2 \subset \mathbb{R}$, the definition

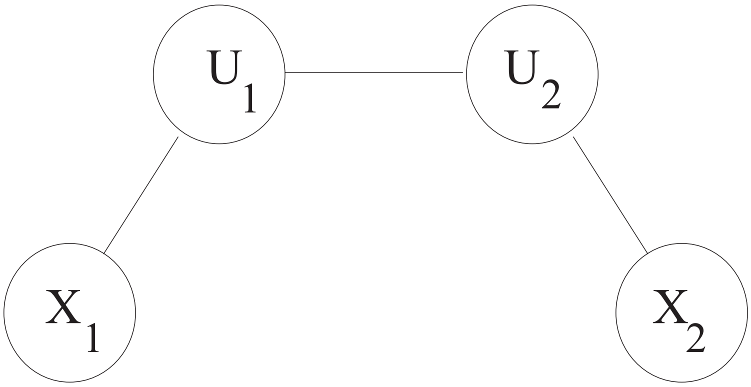

FIG. 7. *Markov tree for coupling*

reduces to the usual definition of a copula. The above definition is important for applications because φ_i might be a physically meaningful function of a set of random variables. The definition could be generalized to allow $\varphi_i(x_i)$ to be a random variable rather than constant. For simulation purposes however it is practical to take deterministic φ_i, as this allows pre-computation of the level sets of φ_i, and hence of the conditional distributions of X_i given U_i. One of the motivations for this approach is that the random quantities may represent physical quantities (for example, temperature, pressure etc). Physical laws, for example the ideal gas law

$$PV = nRT,$$

where P is pressure, V is volume, n is number of moles, T is temperature, and R is the ideal gas constant can be used to give an approximate relationship between variables. Suppose that two vessels of uncertain volume containing an uncertain number of moles of an ideal gas under unknown pressure are placed in the same building. In this case the temperatures of the two vessels would be highly correlated, and one might build a subjective probability model in which the distributions on the quantities P_i, V_i, and n_i for vessel i $(i = 1, 2)$ are coupled via a copula model for the temperatures T_i. The functions φ_i would be

$$T_i = \varphi_i(P_i, V_i, n_i) = \frac{P_i V_i}{n_i R}.$$

We shall also need the notion of a conditional coupling. We suppose that V_D is a probability space and that $d \in V_D$.

DEFINITION 15. *The* $(\varphi_{1|d}, \varphi_{2|d}, c_d)$-family of conditional couplings of families of marginal distributions $(\mu_{1|d}, \mu_{2|d})$ *on the product probability space* $V_1 \times V_2$*, is the family of couplings indexed by* $d \in V_D$ *given by taking the* $(\varphi_{1|d}, \varphi_{2|d}, c_d)$*-coupling of* $\mu_{1|d}$ *and* $\mu_{2|d}$ *for each* d*.*
We say that such a family of conditional couplings is regular *if* $\varphi_{i|d}(x_i)$ *is a measurable function of* (x_i, d) $(i = 1, 2)$*, and if the family of copulae* c_d *is a regular family of conditional probabilities (that is, for all Borel sets* $B \subset [0, 1]^2$*, the mapping* $d \mapsto c_d(B)$ *is measurable).*

The next lemma uses the notation of Definitions 14 and 15. It shows that we really have defined a (family of) couplings, and under what circumstances we can define a probability measure over $V_1 \times V_2 \times V_D$ that has the appropriate marginals.

LEMMA 8.

1. *For any d, the marginal distribution of the $(\varphi_{1|d}, \varphi_{2|d}, c_d)$-coupling measure on V_i is $\mu_{i|d}$, $(i = 1, 2)$.*
2. *Suppose that we are given*
   (a) *joint distributions $\mu_{1,d}$ and $\mu_{2,d}$ on $V_1 \times V_D$ and $V_2 \times V_D$ respectively with the same marginal μ_d on V_D, and*
   (b) *a regular family $(\varphi_{1|d}, \varphi_{2|d}, c_d)$ of conditional couplings.*

   Then there is a joint distribution $\mu_{1,2,d,u_1,u_2}$ on $V_1 \times V_2 \times V_D \times [0,1] \times [0,1]$, such that $\mu_{i,d}$ are marginals $(i = 1, 2)$ and that the induced conditional distribution $\mu_{u_1,u_2|d} = c_d$ for almost all d.

Proof:

1. This follows easily immediately from the remark after Definition 14.
2. Define a random vector (x_i, d) with distribution $\mu_{i,d}$, and then simply define μ_{i,d,u_i} to be the distribution of $(x_i, d, F_{i|d}(\varphi_{i|d}(x_i)))$, where $F_{i|d}$ is the conditional distribution function for $\varphi_{i|d}(x_i)$ given d. We can now form the conditional probabilities $\mu_{i|d,u_i}$ and the marginal μ_d, and then define

$$\mu_{1,2,d,u_1,u_2} = \mu_{1|d,u_1} \mu_{2|d,u_2} c_d \mu_d.$$

□

DEFINITION 16 CANTOR TREE. *A* Cantor tree *on a set of nodes N is a finite set of subsets of N, $\{A_\emptyset, A_1, A_2, A_{11}, A_{12}, A_{21}, A_{22}, \ldots\}$ such that the following properties hold:*

1. *$A_\emptyset = N$.*
2. *(Union property) $A_{i_1 \ldots i_n} = A_{i_1 \ldots i_n 1} \cup A_{i_1 \ldots i_n 2}$, with $A_{i_1 \ldots i_n 1} - A_{i_1 \ldots i_n 2} \neq \emptyset$ and $A_{i_1 \ldots i_n 2} - A_{i_1 \ldots i_n 1} \neq \emptyset$ for all $i_1 \ldots i_n$.*
3. *(Weak intersection property) $D_{i_1 \ldots i_n} := A_{i_1 \ldots i_n 1} \cap A_{i_1 \ldots i_n 2}$ is equal to $A_{i_1 \ldots i_n 1 i_{n+2} \ldots i_m}$, and $A_{i_1 \ldots i_n 2 i'_{n+2} \ldots i'_{m'}}$ for some $i_{n+2} \ldots i_m$ and $i'_{n+2} \ldots i'_{m'}$.*
4. *(Unique decomposition property) If $A_{i_1 \ldots i_n} = A_{j_1 \ldots j_k}$ then for all $t_1 \ldots t_m$,*

$$A_{i_1 \ldots i_n t_1 \ldots t_m} = A_{j_1 \ldots j_k t_1 \ldots t_m}.$$

5. *(Maximal word property) We say that $i_1 \ldots i_n 1$ and $i_1 \ldots i_n 2$ are* maximal *if $D_{i_1 \ldots i_n} = \emptyset$. For any two maximal words $i_1 \ldots i_n$ and $j_1 \ldots j_k$, we have*

$$A_{i_1 \ldots i_n} \cap A_{j_1 \ldots j_k} \neq \emptyset \text{ implies } A_{i_1 \ldots i_n} = A_{j_1 \ldots j_k},$$

   and $\# A_{i_1 \ldots i_n} = 1$.

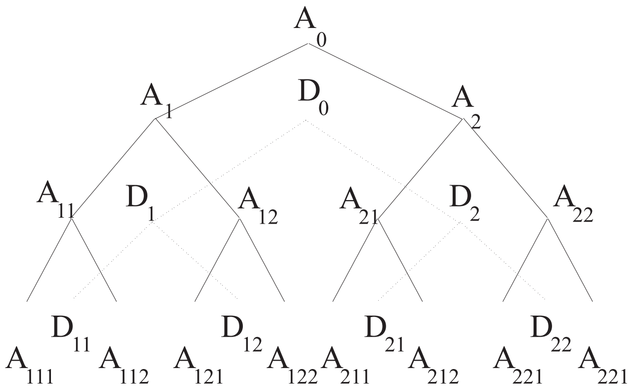

Fig. 8. *Cantor tree*

The name Cantor tree has been chosen because the sets are labeled according to a Cantor set type structure, the binary tree. This is illustrated in Figure 8. In order to make the notation more suggestive concerning the relation between Cantor trees and vines, we introduce the notation $C_{i_1\ldots i_n j} = A_{i_1\ldots i_n j} - D_{i_1\ldots i_n}$ for $j = 1, 2$.

Definition 17 Cantor tree specification. *A* Cantor tree specification *of a multivariate distribution with n variables is a Cantor tree on $N = \{1, \ldots, n\}$ such that the following properties hold:*

1. *(Marginal specification) If $j_1 \ldots j_k$ is maximal then the joint distribution of X_i ($i \in A_{j_1\ldots j_k}$) is specified.*
2. *(Conditional coupling specification) For each $i_1 \ldots i_n$ the conditional coupling of the variables $X_{C_{i_1\ldots i_n 1}}$ and $X_{C_{i_1\ldots i_n 2}}$ given the variables $X_{D_{i_1\ldots i_n}}$ is required to be in a given set of conditional couplings $B_{i_1\ldots i_n}(X_{D_{i_1\ldots i_n}})$.*
3. *(Unique decomposition property) If $A_{i_1\ldots i_n} = A_{j_1\ldots j_k}$ the conditional coupling or marginal specifications are identical.*

Definition 18 Cantor tree dependence. *We say that a distribution F realizes a Cantor tree specification, or exhibits* Cantor tree dependence, *if it satisfies all constraints, that is for all $i_1 \ldots i_n$, the conditional coupling of $C_{i_1\ldots i_n 1}$ and $C_{i_1\ldots i_n 2}$ given $D_{i_1\ldots i_n}$ is a member of the set specified by $B_{i_1\ldots i_n}$, and the marginals of F are those given in the Cantor tree specification.*

Notation: We say that $A_{i_1\ldots i_n}$ is at *level n*. We write $B \leq C$ if $B = A_{i_1\ldots i_n i_{n+1}\ldots i_m}$ and $C = A_{i_1\ldots i_n}$, and say that B *is in the decomposition of* C (note that if $B \leq C$ then $B \subseteq C$ but that the reverse does not have to hold). If $B \leq C$ and $B \neq C$ then we write $B < C$.

We begin by showing that the regular vines of Figures 2 and 3 can be modelled by a Cantor tree.

EXAMPLE 1. *Here we have $N = \{1, 2, 3, 4, 5\}$. The table gives, on each line, a word $*$, followed by the sets A_*, $C_{*,1}$, $C_{*,2}$, and D_*.*

$*$	A_*	$C_{*,1}$	$C_{*,2}$	D_*
$\emptyset$	12345	1	5	234
1	1234	1	4	23
2	2345	5	2	34
11	123	1	3	2
12	234	2	4	3
21	345	3	5	4
22	234	2	4	3
111	12	1	2	$\emptyset$
112	23	2	3	$\emptyset$
121	23	2	3	$\emptyset$
122	34	3	4	$\emptyset$
211	34	3	4	$\emptyset$
212	45	4	5	$\emptyset$
221	23	2	3	$\emptyset$
222	34	3	4	$\emptyset$

The constraints here are precisely the same as those determined by the regular vine in Figure 2.

EXAMPLE 2. *This is an example with $N = \{1, 2, 3, 4, 5\}$.*

$*$	A_*	$C_{*,1}$	$C_{*,2}$	D_*
$\emptyset$	12345	1	5	234
1	1234	1	4	23
2	2345	4	5	23
11	123	1	3	2
12	234	2	4	3
21	234	2	4	3
22	235	2	5	3
111	12	1	2	$\emptyset$
112	23	2	3	$\emptyset$
121	23	2	3	$\emptyset$
122	34	3	4	$\emptyset$
211	23	2	3	$\emptyset$
212	34	3	4	$\emptyset$
221	23	2	3	$\emptyset$
222	35	3	5	$\emptyset$

This corresponds to the vine in Figure 3.

Not all Cantor tree constructions are realizable by regular vines. The point is that the sets $A_{i_1 \ldots i_n 1} - A_{i_1 \ldots i_n 2}$ need not be singletons, as in the next example.

EXAMPLE 3. *This is an example with $N = \{1, 2, 3, 4, 5\}$. A vine corresponding to this example is shown in Figure 9.*

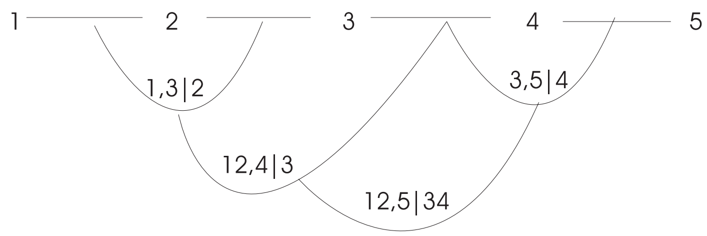

FIG. 9. *Vine representation of Example 3*

$*$	A_*	$C_{*,1}$	$C_{*,2}$	D_*
$\emptyset$	12345	12	5	34
1	1234	12	4	3
2	345	5	3	4
11	123	1	3	2
12	34	3	4	$\emptyset$
21	45	4	5	$\emptyset$
22	34	3	4	$\emptyset$
111	12	1	2	$\emptyset$
112	23	2	3	$\emptyset$

As seems reasonable from the first two examples given above, the constraints determined by a regular vine can always be written in terms of a Cantor tree specification. This will be proven in the next section. Hence Cantor tree specifications are more general than regular vines. We shall show soon however that all Cantor trees can be graphically represented by vines (though not necessarily regular vines). First, however, we prove some results about the existence of Cantor tree dependent distributions.

LEMMA 9. *Suppose distributions $\mu_{A_{i_1\dots i_n 1}}$ and $\mu'_{A_{i_1\dots i_n 2}}$ are given and that the marginals $\mu_{D_{i_1\dots i_n}}$ and $\mu'_{D_{i_1\dots i_n}}$ are equal. Suppose also that a regular family of conditional couplings $(\varphi_{1|d}, \varphi_{2|d}, c_d)$ is given (indexed by the elements d of $D_{i_1\dots i_n}$).*
Then there is a unique distribution $\mu_{A_{i_1\dots i_n}}$ which marginalizes to $\mu_{A_{i_1\dots i_n 1}}$ and $\mu'_{A_{i_1\dots i_n 2}}$ and which is consistent with the family of conditional couplings.

Proof: This follows directly from Lemma 8 by integrating out the variables u_1, u_2. $\square$

THEOREM 6. *Any Cantor tree specification, whose coupling restrictions permit a regular family of couplings for each word $i_1 \dots i_n$, is realised by a Cantor tree dependent distribution over the variables $\{X_i | i \in N\}$.*

Proof: The proof is by induction from the ends of the tree. At any level $i_1 \dots i_n$ in the tree, we assume by induction that the marginals $\mu_{A_{i_1\dots i_n 1}}$ and $\mu'_{A_{i_1\dots i_n 2}}$ are given. By the weak intersection property, the marginal on the

intersection $D_{i_1\ldots i_n}$ has already been calculated earlier in the induction, and by the unique decomposition property the marginals $\mu_{D_{i_1\ldots i_n}}$ and $\mu'_{D_{i_1\ldots i_n}}$ are equal.
The induction argument works whenever $D_{i_1\ldots i_n} \neq \emptyset$. If $D_{i_1\ldots i_n} = \emptyset$ the maximal word property and the unique decomposition property imply a consistent specification. This proves the theorem. □

Remark: We have defined Cantor trees in such a way that the underlying trees are binary, that is, the tree splits in two at each branching point. As a referee pointed out, this is not a necessary requirement. One could easily define a construction with higher order branching. This would involve having several sets $C_{*,1}, \ldots, C_{*,k}$, all with the same set D_*, for any word $*$. The definition of a Cantor tree is adapted in the obvious way.

We now show that there is a simple expression for the information of a distribution realising a Cantor tree specification. Recall that when A_1 and A_2 are specified we use the notation $C_1 = A_1 - A_2$, $C_2 = A_2 - A_1$, and $D = A_1 \cap A_2$.

Lemma 10.

$$\begin{aligned} I(\mu|\prod \mu_i) = I(\mu_{A_1}|\prod_{A_1} \mu_i) + I(\mu_{A_2}|\prod_{A_2} \mu_i) \\ -I(\mu_D|\prod_D \mu_i) + E_D I(\mu_{C_1C_2|D}|\mu_{C_1|D}\mu_{C_2|D}). \end{aligned}$$

This follows in the same way as Lemma 7(5).

Theorem 7 Information Decomposition Theorem. *For a Cantor tree dependent distribution μ, we have*

$$I(\mu|\prod \mu_i) = \sum_{\{A_{i_1\ldots i_n}, D_{i_1\ldots i_n}\}} E_{D_{i_1\ldots i_n}} I(\mu_{C_{i_1\ldots i_n1}C_{i_1\ldots i_n2}|D_{i_1\ldots i_n}}|\mu_{C_{i_1\ldots i_n1}|D_{i_1\ldots i_n}}\mu_{C_{i_1\ldots i_n2}|D_{i_1\ldots i_n}}).$$

The index of the summation sign says that the terms in the summation occur once for each $\{A_{i_1\ldots i_n}, D_{i_1\ldots i_n}\}$, that is, the collection of pairs $A_{j_1\ldots j_k}$, $D_{j_1\ldots j_k}$ with $A_{i_1\ldots i_n} = A_{j_1\ldots j_k}$ and $D_{i_1\ldots i_n} = D_{j_1\ldots j_k}$ contributes just one term to the summation. When the conditioning set $D_{i_1\ldots i_n}$ is empty then the conditional information term is constant and the expectation operation gives (by convention) that constant value.
Proof: Consider first the expression obtained by applying Lemma 10 repeatedly from the top of the tree. We have

$$\begin{aligned} I(\mu|\prod \mu_i) &= I(\mu_{A_1}|\prod_{A_1} \mu_i) + I(\mu_{A_2}|\prod_{A_2} \mu_i) \\ &\quad -I(\mu_D|\prod_D \mu_i) + E_D I(\mu_{C_1C_2|D}|\mu_{C_1|D}\mu_{C_2|D}) \\ &= I(\mu_{A_{11}}|\prod_{A_{11}} \mu_i) + I(\mu_{A_{12}}|\prod_{A_{12}} \mu_i) \\ &\quad -I(\mu_{D_1}|\prod_{D_1} \mu_i) + E_{D_1} I(\mu_{C_{11}C_{12}|D_1}|\mu_{C_{11}|D_1}\mu_{C_{12}|D_1}) \end{aligned}$$

$$+I(\mu_{A_{21}}|\prod_{A_{21}}\mu_i) + I(\mu_{A_{22}}|\prod_{A_{22}}\mu_i)$$

$$-I(\mu_{D_2}|\prod_{D_2}\mu_i) + E_{D_2}I(\mu_{C_{21}C_{22}|D_2}|\mu_{C_{21}|D_2}\mu_{C_{22}|D_2})$$

$$-I(\mu_D|\prod_D\mu_i) + E_D I(\mu_{C_1C_2|D}|\mu_{C_1|D}\mu_{C_2|D})$$

We expand in this way until we reach terms for which $D_{i_1\ldots i_n} = \emptyset$.
In carrying out this expansion we obtain negative terms of the form

$$-I(\mu_{D_{i_1\ldots i_n}}|\prod_{D_{i_1\ldots i_n}}\mu_i).$$

The weak intersection property says however that every non-empty $D_{i_1\ldots i_n}$ is equal to two $A_{j_1\ldots j_k}$ later in the expansion. Hence the $-I(\mu_{D_{i_1\ldots i_n}}|\prod_{D_{i_1\ldots i_n}}\mu_i)$ term is added to a

$$2I(\mu_{D_{i_1\ldots i_n}}|\prod_{D_{i_1\ldots i_n}}\mu_i)$$

term arising later in the expansion of the summation.
We now claim that each term arising in the expansion has multiplicity equal to one. Suppose we have two words $i_1\ldots i_n$ and $j_1\ldots j_k$ with $A_{i_1\ldots i_n} = A_{j_1\ldots j_k}$ and $D_{i_1\ldots i_n} = D_{j_1\ldots j_k}$. Write $i_1\ldots i_m$ for the longest words common to $i_1\ldots i_n$ and $j_1\ldots j_k$, that is, $i_\ell = j_\ell$ for $\ell = 1,\ldots,m$ and (without loss of generality) $i_{m+1} = 1 \neq 2 = j_{m+1}$. Then $A_{i_1\ldots i_n} \subseteq A_{i_1\ldots i_m 1}$ and $A_{j_1\ldots j_k} \subseteq A_{i_1\ldots i_m 2}$. Hence $A_{i_1\ldots i_n} \subseteq D_{i_1\ldots i_m}$, and, by Lemma 11 below, $A_{i_1\ldots i_n}$ is in the decomposition of $D_{i_1\ldots i_m}$. The same holds for $A_{j_1\ldots j_k}$. This shows that one of the two terms in the summation arising from $A_{i_1\ldots i_n}$ and $A_{j_1\ldots j_k}$ will be cancelled by a negative term occurring in the expansion of the $-I(\mu_{D_{i_1\ldots i_m}}|\prod_{D_{i_1\ldots i_m}}\mu_i)$ term.
Note that if there are three words with identical $A_{i_1\ldots i_n}$ then they cannot all share a common longest word, so the argument of the previous paragraph can be used inductively to show that the extra terms are cancelled out.
This proves the theorem. □

LEMMA 11. *Suppose $B < A_{i_1\ldots i_n}$ and $B \subseteq D_{i_1\ldots i_n}$. Then $B \leq D_{i_1\ldots i_n}$.*

Proof: The statement will be proved by backwards induction on n. When $B = D_{i_1\ldots i_n}$ the lemma is obvious, so we assume from now on that $B \neq D_{i_1\ldots i_n}$.
When $i_1\ldots i_n$ is a maximal word or $i_1\ldots i_n i_{n+1}$ is maximal then the statement holds trivially.
Now take a general n. For ease of notation we denote $D_{i_1\ldots i_n}$ simply by D. Since we have $B < A_{i_1\ldots i_n}$ and $D < A_{i_1\ldots i_n}$, there are words $i_1\ldots i_n b_{n+1}\ldots b_{n'}$ and $i_1\ldots i_n d_{n+1}\ldots d_{n''}$ such that

$$B = A_{i_1\ldots i_n b_{n+1}\ldots b_{n'}} \quad \text{and} \quad D = A_{i_1\ldots i_n d_{n+1}\ldots d_{n''}}.$$

Amongst all such possible words choose a pair with the longest common starting word $i_1\ldots i_n i_{n+1}\ldots i_m$. Clearly $m \geq n$. In fact, $m \geq n+1$ since $B \leq A_{i_1\ldots i_n i_{n+1}}$ and by the properties of a Cantor tree, $D \leq A_{i_1\ldots i_n i_{n+1}}$.

We now have that $B \leq A_{i_1 \ldots i_m}$ and $D \leq A_{i_1 \ldots i_m}$. Since $B \subset D$ we must have $B < A_{i_1 \ldots i_m}$. Assume for a contradiction that $B \not\leq D$, then also $D < A_{i_1 \ldots i_m}$. The maximality of m then implies that (without loss of generality)

$$B \leq A_{i_1 \ldots i_m 1}, \quad D \not\leq A_{i_1 \ldots i_m 1}$$

and

$$D \leq A_{i_1 \ldots i_m 2}, \quad B \not\leq A_{i_1 \ldots i_m 2}.$$

But now since $B \subset D$ we must have $B \subset D_{i_1 \ldots i_m}$ and by the induction hypothesis $B \leq D_{i_1 \ldots i_m}$ so that also $B \leq A_{i_1 \ldots i_m 2}$. This contradicts maximality of m. □

COROLLARY 2. *If, for all $i_1 \ldots i_n$ and almost all $d \in D_{i_1 \ldots i_n}$, the conditional distribution $\mu_{C_{i_1 \ldots i_n 1} C_{i_1 \ldots i_n 2}|d}$ has minimum information (relative to the independent joint distribution), then the Cantor tree dependent distribution has minimum information amongst all Cantor tree dependent distributions satisfying the Cantor tree specification.*

6. Vine representations and Cantor trees The purpose of this section is to show that *regular* vines can be represented by Cantor trees, and that Cantor trees can be represented by vines.
We first show:

THEOREM 8. *Any regular vine dependent distribution can also be represented by a Cantor tree dependent distribution.*

Proof: It is enough to show that any regular vine constraints can be encoded by Cantor tree constraints.
Let $\mathcal{V}$ be a regular vine. We construct a Cantor tree corresponding to $\mathcal{V}$ by defining a mapping ϕ from binary words to nodes in the vine.
We set $\phi(\emptyset)$ to equal the single node in T_n. The map ϕ is defined further by induction. Suppose that ϕ is defined on all binary words of length less than m. Let w be a word of length $m-1$, with $\phi(w) = e = \{j,k\}$. We define $\phi(w1) = j$ and $\phi(w2) = k$ arbitrarily.
Now, for any binary word w we define $A_w = A_{\phi(w)}$, and claim that the collection $\{A_w\}$ so formed is a Cantor tree.
The union property follows because when $e = \{j,k\}$, we have $A_e = A_j \cup A_k$. The weak intersection propery follows from the proof of Lemma 5. The unique decomposition property follows from Lemma 6. When w is maximal, A_w is a singleton, so that the maximal word property holds trivially.
It is now easy to see that this Cantor tree specification is the same as the regular vine specification, and the theorem follows. □

Remark: Short words correspond to nodes in high level trees in the regular vine, while long words correspond to nodes in low level trees. This arises because a Cantor tree is a "top-down" construction, while a vine is a "bottom-up" construction.

This result implies that the proof of existence of Cantor tree dependent distributions given in the last section applies also to regular vine dependent distributions.

We now show that Cantor tree specifications can also be represented by vines. As an example, Figure 9 shows the vine representation of the Cantor tree specification given in Example 3. Checking the formal definition of a vine, we see that for this example one can choose $m = 4$ and further:

1) $T_1 = (E_1, N_1)$ with
$$N_1 = \{1, 2, 3, 4, 5\},$$
and $E_1 = \{\{1, 2\}, \{2, 3\}\}$.

2) $T_2 = (E_2, N_2)$ with
$$\begin{aligned} N_2 &= \{\{1, 2\}, \{2, 3\}, 3, 4, 5\} \\ &\subset E_1 \cup N_1, \end{aligned}$$
and
$$E_2 = \{\{\{1, 2\}, \{2, 3\}\}, \{3, 4\}, \{4, 5\}\}.$$

3) $T_3 = (E_3, N_3)$ with
$$\begin{aligned} N_3 &= \{\{\{1, 2\}, \{2, 3\}\}, \{3, 4\}, \{4, 5\}\} \\ &\subset E_2 \cup E_1 \cup N_1, \end{aligned}$$
and
$$E_3 = \{\{\{\{1, 2\}, \{2, 3\}\}, \{3, 4\}\}, \{\{3, 4\}, \{4, 5\}\}.$$

4) $T_4 = (E_4, N_4)$ with
$$N_4 = E_3 \subset E_1 \cup N_1,$$
and
$$E_4 = \{\{\{\{\{1, 2\}, \{2, 3\}\}, \{3, 4\}\}, \{\{3, 4\}, \{4, 5\}\}\}.$$

More generally, one can always construct a vine representation of a Cantor tree specification in this way, as will be shown below (the main problem is to show that at each level one has a tree). A vine is a useful way of representing such a specification as it guarantees that the union and the unique decomposition properties hold. The only property that does not have to hold for a vine is the weak intersection property. The vine in Figure 6 does not have the weak intersection property.

THEOREM 9. *Any Cantor tree specification has a corresponding vine representation.*

Proof: Let m be the maximum length of a maximal word. Define $T_1 = \{N_1, E_1\}$, where $N_1 = N$ and $e = \{j, k\} \in E_1$ if and only if for some word w of length $m - 1$,
$$A_{w1} = \{j\}, \quad \text{and} \quad A_{w2} = \{k\}.$$
More generally, $e = \{j, k\} \in E_\ell$ if and only if $e \notin E_{\ell-1} \cup \ldots \cup E_1$ and for some word w of length $m - \ell$, A_{w1} equals the complete union of j and A_{w2} equals the complete union of k. This inductively defines the pairs $T_i = (N_i, E_i)$

$(i = 1, \ldots, m)$. However, it remains to be shown that these are trees, that is, that there are no cycles.
Suppose for a contradiction there is no cycle in $T_m, \ldots, T_{\ell+1}$, and there is a cycle in T_ℓ. Without loss of generality there are nodes $A_{i_1 \ldots i_n}$ and $A_{j_1 \ldots j_k}$ on the cycle with $i_1 = 1$, $j_1 = 2$ and such that $A_{i_1 \ldots i_n} \not< A_2$, $A_{j_1 \ldots j_k} \not< A_1$. Then there must be at least two nodes in the cycle that are subsets of D. Then by Lemma 11, they are in the decomposition of D and hence also in the decomposition of A_1 and of A_2. There must also be a path joining these two nodes by nodes in the decomposition of D. Hence there is a cycle containing the two nodes with one of the two arcs joining the two nodes made up of nodes just in the decomposition of A_1 (say), and the other arc of the cycle is made up of nodes in the decomposition of D and thus also of A_2. But then the nodes in $T_{\ell+1}$ which are the edges of the cycle form a cycle in $T_{\ell+1}$. This contradicts the assumption that ℓ was the largest integer for which T_ℓ contains a cycle. □

7. Rank and partial correlation specifications

In this section we discuss vine constructions in which we specify correlations on each vine branch.

7.1. *Partial correlation specifications*

We first recall the definition and interpretation of partial correlation.

Definition 19 partial correlation. *Let $X_1, \ldots, X_n$ be random variables. The partial correlation of X_1 and X_2 given $X_3, \ldots, X_n$ is*

$$\rho_{12|3,\ldots,n} = \frac{\rho_{12|4\ldots n} - \rho_{13|4\ldots n}\rho_{23|4\ldots n}}{((1-\rho^2_{13|4\ldots n})(1-\rho^2_{23|4\ldots n}))^{\frac{1}{2}}}.$$

If $X_1, \ldots, X_n$ follow a joint normal distribution with variance covariance matrix of full rank, then partial correlations correspond to conditional correlations. In general, all partial correlations can be computed from the correlations by iterating the above equation. Here we shall reverse the process, and for example use a regular vine to specify partial correlations in order to obtain a correlation matrix for the joint normal distribution.

Definition 20 partial correlation vine specification. *If $\mathcal{V}$ is a regular vine on n elements, and $e \in E_i$, then a* complete partial correlation specification *is a regular vine with a partial correlation p_e specified for each edge e. A distribution satisfies the complete partial correlation specification if, for any edge $e = \{j,k\}$ in the vine, the partial correlation of the variables in $C_{e,j}$ and $C_{e,k}$ given the variables in D_e is equal to p_e.*
A complete normal partial correlation specification *is a special case of a regular vine specification, denoted triple $(\underline{F}, \mathcal{V}, \rho)$, satisfying the following condition: For every e and vector of values d taken by the variables in D_e, the set $B_e(d)$ just contains the single normal copula with correlation ρ_e (which is constant in d).*

Remark: We have defined a partial correlation specification without reference to a family of copulae as, in general, the partial correlation is not a property of a copula. For the bivariate normal distribution, however, this is the case.

As remarked above, partial correlation is just equal to conditional correlation for joint normal variables. The meaning of partial correlation for non-normal variables is less clear. We quote Kendall and Stuart [15](p335): "In other cases [i.e. non-normal], we must make due allowance for observed heteroscedasticity in our interpretations: the partial regression coefficients are then, perhaps, best regarded as average relationships over all possible values of the fixed variates."
If $\mathcal{V}$ is a regular vine over n elements, a partial correlation specification stipulates partial correlations for each edge in the vine. There are $\binom{n}{2}$ edges in total, hence the number of partial correlations specified is equal to the number of pairs of variables, and hence to the number of ρ_{ij}. Whereas the ρ_{ij} must generate a positive definite matrix, the partial correlations of a regular vine specification may be chosen arbitrarily from the interval $(-1, 1)$.
The following lemma summarizes some well-known facts about conditional normal distributions (see for example [23]).

Lemma 12. *Let $X_1, \ldots, X_n$ have a joint normal distribution with mean vector $(\mu_1, \ldots, \mu_n)'$ and covariance matrix Σ. Write Σ_A for the principal submatrix built from rows 1 and 2 of Σ, etc so that*

$$\Sigma = \begin{pmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_B \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}.$$

Then the conditional distribution of $(X_1, X_2)'$ given $(X_3, \ldots, X_n)' = x_B$ is normal with mean $\mu_A + \Sigma_{AB}\Sigma_B^{-1}(x_B - \mu_B)$ and covariance matrix

$$\Sigma_{12|3\ldots n} = \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA}. \tag{7.7}$$

Writing $\sigma_{ij|3\ldots n}$ for the i, j-element of Σ_{AA}, the partial correlation satisfies

$$\rho_{12|3\ldots n} = \frac{\sigma_{12|3\ldots n}}{\sqrt{\sigma_{11|3\ldots n}\sigma_{22|3\ldots n}}}.$$

Hence, for the joint normal distribution, the partial correlation is equal to the conditional product moment correlation. The partial correlation 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 $X_3, \ldots, X_n$.
The next lemma will be used to couple normal distributions together. The symbol $< v, w >$ denotes the usual Euclidean inner product of two vectors. The proof works by embedding the first set of n-dimensional vectors in $\mathbb{R} \times \mathbb{R}^{n-1} \subset \mathbb{R} \times \mathbb{R}^{n-1} \times \mathbb{R}$, and the second set in $\mathbb{R}^{n-1} \times \mathbb{R} \subset \mathbb{R} \times \mathbb{R}^{n-1} \times \mathbb{R}$.

Lemma 13. *Let $v_1, \ldots, v_{n-1}$ and $u_2, \ldots, u_n$ be two sets of linearly independent vectors of unit length in $\mathbb{R}^{n-1}$. Suppose that*

$$< v_i, v_j > = < u_i, u_j > \quad \text{for } i, j = 2, \ldots, n-1.$$

Then given $\alpha \in (-1, 1)$ we can find a linearly independent set of vectors of unit length $w_1, \ldots, w_n$ in $\mathbb{R}^n$ such that

1. $< w_i, w_j >=< v_i, v_j >$ *for* $i = 1, \ldots, n-1$
2. $< w_i, w_j >=< u_i, u_j >$ *for* $i = 2, \ldots, n$
3. $< w_1', w_n' >= \alpha$, *where* w_1' *and* w_n' *denote the normalized orthogonal projections of* w_1 *and* w_n *onto the orthogonal complement of the space spanned by* $w_2, \ldots, w_{n-1}$.

The corollary to this lemma follows directly using the interpretation of a positive definite matrix as the matrix of inner products of a set of linearly independent vectors.

Corollary 3. *Suppose that* $(X_1, \ldots, X_{n-1})$ *and* $(Y_2, \ldots, Y_n)$ *are two multivariate normal vectors, and that* $(X_2, \ldots, X_{n-1})$ *and* $(Y_2, \ldots, Y_{n-1})$ *have the same distribution. Then for any* $-1 < \alpha < 1$, *there exists a multivariate normal vector* $(Z_1, \ldots, Z_n)$ *so that*

1. $(Z_1, \ldots, Z_{n-1})$ *has the distribution of* $(X_1, \ldots, X_{n-1})$,
2. $(Z_2, \ldots, Z_n)$ *has the distribution of* $(X_2, \ldots, X_n)$, *and*
3. *the partial correlation of* Z_1 *and* Z_n *given* $(Z_2, \ldots, Z_{n-1})$ *is* α.

We now show how the notion of a regular vine can be used to construct a joint normal distribution.

Theorem 10. *Given any complete partial correlation vine specification for standard normal random variables* $X_1, \ldots, X_n$, *there is a unique joint normal distribution for* $X_1, \ldots, X_n$ *satisfying all the partial correlation specifications.*

Proof: We use the Cantor tree representation of the regular vine. The proof is by induction in the Cantor tree. Clearly any two normal variables can be given a unique joint normal distribution with the product moment rank correlation strictly between -1 and 1.
Suppose that for any binary word w longer than length k, the variables in A_w can be given a unique joint normal distribution consistent with the partial correlations given in the vine. Consider now a binary word v of length $k-1$. Since the vine is regular, we can write A_v as a disjoint union

$$A_v = C_{v1} \cup C_{v2} \cup D_v,$$

where C_{v1} and C_{v2} both contain just one element. The corresponding node in the regular vine specifies the partial correlation of C_{v1} and C_{v2} given D_v. By the induction hypothesis there is a unique joint normal distribution on the elements of A_{v1} and similarly a unique joint normal distribution on the elements of A_{v2}, all satisfying the vine constraints on these elements. Furthermore, the distributions must both marginalize to the same joint normal distribution on D_v. Hence we are in the situation covered by Corollary 3, and we can conclude that the variables of A_v can be given a joint normal distribution which marginalizes to the distributions we had over A_{v1} and A_{v2}, and which has the partial correlation coefficient for C_{v1} and C_{v2} given D_v that was given in the specification of the vine. □

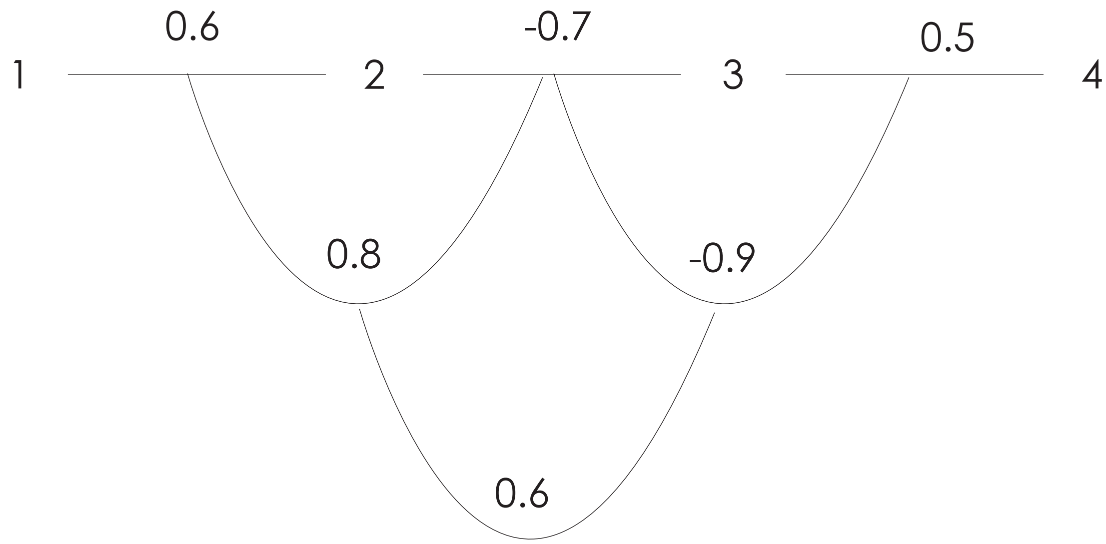

FIG. 10. *Partial correlation vine*

COROLLARY 4. *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.*

We note that unconditional correlations can easily be calculated inductively by using Equation 7.7. This is demonstrated in the following example.

EXAMPLE 4. *Consider the vine in Figure 10. We consider the subvine consisting of nodes 1,2 and 3. Writing the correlation matrix with the variables ordered as 1,3,2, we wish to find a product moment correlation ρ_{13} such that the correlation matrix*

$$\begin{pmatrix} 1 & \rho_{13} & 0.6 \\ \rho_{13} & 1 & -0.7 \\ 0.6 & -0.7 & 1 \end{pmatrix}$$

has the required partial correlation. We apply Equation 7.7 with

$$\Sigma_B = (1), \quad \Sigma_A = \begin{pmatrix} 1 & \rho_{13} \\ \rho_{13} & 1 \end{pmatrix}, \quad \Sigma_{AB} = \begin{pmatrix} 0.6 \\ -0.7 \end{pmatrix}, \quad \Sigma_{13|2} = \begin{pmatrix} \sigma_{1|2}^2 & 0.8\sigma_{1|2}\sigma_{3|2} \\ 0.8\sigma_{1|2}\sigma_{3|2} & \sigma_{3|2}^2 \end{pmatrix}.$$

This gives $\sigma_{1|2} = 0.8$, $\sigma_{3|2} = 0.7141$, and

$$\rho_{13} = 0.8\sigma_{1|2}\sigma_{3|2} - 0.42 = 0.0371.$$

Using the same method for the subvine with nodes 2,3, and 4, we easily calculate that the unconditional correlation $\rho_{24} = -0.9066$. In the same way we find that $\rho_{14} = -0.5559$. Hence the full (unconditional) product-moment *correlation matrix for variables 1,2,3, and 4 is*

$$\begin{pmatrix} 1 & 0.6 & 0.0371 & -0.5559 \\ 0.6 & 1 & -0.7 & -0.9066 \\ 0.0371 & -0.7 & 1 & 0.5 \\ -0.5559 & -0.9066 & 0.5 & 1 \end{pmatrix}.$$

Remark: As this example shows, for the *standard vine* on n elements (of which Figures 2 and 10 are examples) in which each tree is linear (that is, there are no nodes of degree higher than 2), the partial correlations can be conveniently written in a symmetric matrix in which the ijth entry $(i < j)$ gives the partial correlation of $ij|i+1,\ldots,j-1$. This matrix, for which all off-diagonal elements of the upper triangle may take arbitrary values between -1 and 1, gives a convenient alternative matrix parameterization of the multivariate normal correlation matrix.

The partial correlations in a vine specify the complete correlation matrix, even with no assumptions of joint normality. This is stated in the following result which may be proved by induction using the formula for partial correlation given in Definition 19.

THEOREM 11. *Let $(X_1,\ldots,X_n)$ and $(Y_1,\ldots,Y_n)$ be vectors of random variables satisfying the same partial correlation vine specification. Then for $i \neq j$,*

$$\rho(X_i, X_j) = \rho(Y_i, Y_j).$$

Our notion of a partial correlation vine specification generalizes a construction of Joe [13] who, in our terminology, defined a partial correlation specification on a standard vine.

7.2. *Rank correlation specifications*

DEFINITION 21 RANK CORRELATION SPECIFICATION. *If $\mathcal{V}$ is a regular vine on n elements, then a complete conditional rank correlation specification is a triple $(\underline{F}, \mathcal{V}, r)$ so that for every e and vector of values d taken by the variables in D_e, every copula in the set $B_e(d)$ has conditional rank correlation $r_e(d)$, $(|r_e(d)| \leq 1)$.*

In Proposition 1 below we show that if $r_e(d)$ is a Borel measurable function of d then the conditional copula family formed by taking the minimal information copula with given rank correlation for a.e. d is a regular conditional probability family.
We now turn to rank correlation specifications.

PROPOSITION 1. *Suppose that X_1, X_2 are random variables, and that X_D is a vector of random variables. Suppose further that the joint distributions of (X_1, X_D) and (X_2, X_D) are given, and that the function*

$$X_D \mapsto r_{X_D}(X_1, X_2)$$

is measurable. Then the conditional copula family formed by taking the minimal information copula with given rank correlation for a.e. X_D is a regular conditional probability family.

Proof: The density function of the minimal rank correlation at any given point varies continuously as function of the rank correlation [22]. Hence for any Borel set B, the minimal information measure of B is a continuous function of the rank correlation. Then the minimal information measure of B is a measurable function of X_D. □

THEOREM 12. *Suppose that we are given a rank tree specification for a regular vine for which the conditional rank correlation functions are all measurable, and the marginals have no atoms. If we take the minimal information copula given the required conditional rank correlation everywhere, then this gives the distribution that has minimal information with respect to the independent distribution with the same marginals.*

Proof: Note first that information is invariant under bi-measurable bijections. Hence, whenever F and G are the distribution functions of continuous random variables X and Y, the information of the copula for X and Y (with respect to the uniform copula) equals that of the joint distribution of X and Y with respect to the independent distribution with the same marginals. It is easy to see that all marginal distributions constructed using minimal information copulae with given rank correlation are continuous. The result now follows from Theorems 7, 8 and Proposition 1.□

8. Conclusions Conditional rank correlation vine specifications can be sampled on the fly, and the minimum information distribution consistent with a rank correlation specification is easily sampled using bivariate minimum information copulae. Moreover, a user specifies such a distribution by specifying $\binom{n}{2}$ numbers in $[-1, 1]$ which needn't satisfy any additional constraint. In the minimum information realisation, a conditional rank correlation of zero between two variables means that the variables are conditionally independent. From a simulation point of view conditional rank correlation specifications are attractive ways to specify high dimensional joint distributions.
One of the more common ways to define a multivariate distribution is to transform each of the variables to univariate normal, and then to take the multivariate normal distribution to couple the variables. The disadvantage of this procedure is that the conditional rank correlations of the variables are always constant (reflecting the constancy of the conditional product moment correlation for the multivariate normal). With vines it is possible to define non-constant conditional rank correlations, and therefore to generate a much wider class of multivariate distributions.

9. Acknowledgements The authors would like to thank I. Meilijson, J. Misiewicz, T. Terlaky, the referees and the editors for helpful comments.

REFERENCES

[1] Arnold, B. C., Castillo, E., Sarabia, J. M., *Conditional specification of statistical models.* Springer Series in Statistics. Springer-Verlag, New York, 1999.

[2] Cooke R.M., Keane M.S. and Meeuwissen A.M.H., *User's Manual for RIAN: Computerized Risk Assessment*, Estec 1233, Delft University of Technology, The Netherlands, (1990).

[3] Cooke R.M., Meeuwissen A.M.H. and Preyssl C., *Modularizing Fault Tree Uncertainty Analysis: The Treatment of Dependent Information Sources*, Probabilistic Safety Assessment and Management Ed. G. Apostolakis, Elsevier Science Publishing Co, (1991).

[4] Cooke, R.M., *UNICORN: Methods and Code for Uncertainty Analysis* AEA Technologies, Warrington, (1995).

[5] Cooke, R.M., "Markov and Entropy Properties of Tree- and Vine-Dependent Variables", *Proceedings of the ASA Section on Bayesian Statistical Science*, 1997.

[6] Cuadras C.M., "Probability Distributions with Given Multivariate Marginals and Given Dependence Structure", *J. of Multiv. Analysis*, vol. 42, pp.51-66, (1992).

[7] Genest, C., Quesada M., José J., Rodriguez L., José A., "De l'impossibilite de construire des lois a marges multidimensionnelles donnees a partir de copules." *C. R. Acad. Sci. Paris Sér. I Math.* 320 (1995), no. 6, 723–726.

[8] Guiaşu S., *Information Theory with Applications*, McGraw-Hill, New York, (1977).

[9] O'Hagan, A., *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*, Edward Arnold, London, 1994.

[10] Iman, R., Helton J. and Campbell, J., "An approach to sensitivity analysis of computer models: Parts I and II" *J. of Quality Technology*, 13 (4), (1981).

[11] Iman, R., and Conover, W., "A distribution-free approach to inducing rank correlation among input variables", *Communications in Statistics - Simulation and Computation*, 11 (3) 311-334, (1982).

[12] Jaynes, E.T., *Papers on Probability, Statistics and Statistical Physics*, R.D. Rosenkrantz (ed.), D. Reidel Publ., Dordrecht, 1983.

[13] Joe, H., "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters", in *Distributions with fixed marginals and related topics*, L. Rüschendorf, B. Schweizer and M.D. Taylor (eds), IMS Lecture Notes Monograph Series, **28**, 120-141, 1996.

[14] Joe, H., *Multivariate models and dependence*, Chapman and Hall, 1997.

[15] Kendall, M.G. and Stuart, A. *The advanced theory of statistics. Vol. 2. Inference and relationship Edition* 2nd ed., Griffin, London 1967.

[16] Kraan, B., *Ph.D Thesis*, Delft University of Technology, Delft, The Netherlands, 2001.

[17] Kullback S., *Information Theory and Statistics*, Wiley, (1959).

[18] Li, H., Scarsini, M., Shaked, M., "Linkages: a tool for the construction of multivariate distributions with given non-overlapping marginals", *J. Multivariate Anal.*, **56**, 20-41, 1996

[19] Li, H., Scarsini, M., Shaked, M., "Dynamic linkages for multivariate distributions with given non-overlapping multivariate marginals", *J. Multivariate Anal.*, **68**, 54-77, 1999.

[20] Meeuwissen A.M.H., *Dependent Random Variables in Uncertainty Analysis*, Ph.D. Thesis, TU Delft, The Netherlands, (1993).

[21] Meeuwissen A.M.H. and Cooke R.M., *Tree dependent random variables* Report 94-28, Department of Mathematics, Delft University of Technology, Delft, The Netherlands, (1994).

[22] Meeuwissen A.M.H. and Bedford, T.J., "Minimal informative distributions with given rank correlation for use in uncertainty analysis", *J. Stat Comp and Simul*, 57, nos 1-4, 143-175, (1997).

[23] Muirhead R.J., *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[24] Oliver R.M. and Smith J.Q. (eds), *Influence Diagrams, Belief Nets and Decision Analysis*, Wiley, Chichester, (1990).