American Statistical Association

1997
Proceedings

of the

Section on Bayesian Statistical Science

Papers presented at the
Annual Meeting of the American Statistical Association,
Anaheim, California, August 10–14, 1997,

The International Society for Bayesian Analysis
4th World Meeting
University of Cape Town, South Africa, December 17–20, 1996

The International Society for Bayesian Analysis
5th World Meeting
Istanbul, Turkey, August 16–18, 1997

all under the sponsorship of the
Section on Bayesian Statistical Science

American Statistical Association ■ 1429 Duke Street, Alexandria, VA 22314

ASA 1997 Proc. Section Bayesian Stat. Sci.

Markov and Entropy Properties of Tree- and Vine-Dependent Variables

Roger M. Cooke

Delft University of Technology
Faculty of Mathematics and Informatics
Mekelweg 4, 2628 CD Delft, The Netherlands

Keywords: correlation, dependence, entropy, multivariate probability distribution, Monte-Carlo simulation, tree dependence, Markov tree.

Mathematics Subject Classification: 62E25; 60H05, 90B25, 94A17.

Abstract:

Bivariate tree specifications are introduced to handle the situation of "partially known" joint probability distributions that arise in Monte Carlo simulation studies. Alternative approaches are found in (Iman et al 1981, Iman and Conover 1982). Bivariate tree specifications are distinguished by the fact that they have Markov tree dependent realizations. These Markov tree dependent realizations have a maximal entropy property, are readily sampled, and are very "smooth". This smoothness can be used to reduce the calculational burden in uncertainty analysis situations as has been shown in e.g. Cooke, Meeuwissen & Preyssl (1991), Meeuwissen (1993) and Meeuwissen & Cooke (1994). Algorithms for the generation of samples of tree dependent random variables have been implemented in computer programs by Cooke, Keane & Meeuwissen (1990) and Cooke (1995), and described in Meeuwissen & Bedford (1997).

Sections 1 through 4 collect results for bivariate tree specifications. Section 5 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-1} are the nodes of T_i. Maximal entropy results show how complicated conditional independence properties can be obtained from vine specifications in combination with entropy maximization. Sampling from maximal entropy distributions given marginal and (conditional) rank correlations specified on a vine can be just as fast as independent sampling.

1. Definitions and Preliminaries

We consider continuous invertible 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 $(1 \leq i \leq n)$ and the bivariate marginal distribution functions are written F_{ij} $(1 \leq i \neq j \leq n)$. The $F_{i|j}$ denotes the distribution of variable i conditional on j. If f is the density of F, then the same subscript conventions apply. Whenever we use the relative information integral, the absolute continuity condition mentioned below is assumed to hold.

Definition 1 (relative information)
If f and g are densities with f absolutely continuous with respect to g then the relative information $I(f|g)$ of density f with respect to g is

$$I(f|g) = \int f(x) \log(\frac{f(x)}{g(x)})\, dx \ .$$

Properties of $I(f|g)$ are that $I(f|g) \geq 0$ and $I(f|g) = 0 \Leftrightarrow f = g$. $I(f|g)$ can be interpreted as measuring the degree of "uniformness" of f (with respect to g). See e.g. Kullback (1959) and Guiaşu (1977).

Definition 2 (rank or Pearson correlation)
The rank correlation $\rho_r(X, Y)$ of two random variables X and Y with a joint probability distribution $F_{X,Y}$ and marginal probability distributions F_X and F_Y respectively, is given by

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y)) \ .$$

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

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

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

- 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.

Definition 3 (tree)
A tree $T = \{N, E\}$ is an acyclic graph, where N is a non empty set (nodes) and E (edges) is a set of unordered pairs of nodes.

Definition 4 (rank correlation tree specfication)

(F, T, t) is an n-dimensional rank correlation tree specification if:

1. *$F = (F_1, \ldots, F_n)$ 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 $t = \{t_{ij} | t_{ij} \in [-1, 1], \{i, j\} \in E, t_{ij} = t_{ji}, t_{ii} = 1\}$.*

Definition 5 (bivarate tree specification)
(F,T,B) is an n-dimensional bivariate tree specification if:

1. *$F = (F_1, \ldots, F_n)$ is a vector of one-dimensional distribution functions,*
2. *T is a tree with nodes $N = \{1, \ldots, n\}$ and edges E*
3. *$B = \{B(i, j); \{i, j\}b \in E\}$; where $B(ij)$ is a subset of the class of distribution functions with marginals F_i and F_j.*

Definition 6 (tree dependence)
(i) A multivariate probability distribution G on $\mathbb{R}^n$ satisfies, or realizes, a bivariate tree specification (F, T, B) if the marginal distributions G_i of G equal F_i $(1 \le i \le n)$ and if for $\{i, j\} \in E$ the bivariate distributions G_{ij} of G are elements of $B(ij)$.
(ii) G has tree dependence order M for T if $\{i, k_1\}, \ldots, \{k_m, j\} \in E$ implies that X_i and X_j are conditionally indepedent given any M of k_ℓ, $1 \le \ell \le m$; and if X_i and X_j are independent if there are no such $k_1, \ldots, k_m$ $(i, j \in N)$.
(iii) G has Markov tree dependence for T if G has tree dependence order M for all M for all $M \in N$,

The following results are proved in (Meeuwissen and Cooke 1994). The first is similar to results with influence diagrams (Oliver and Smith, 1990), the second uses a construction of (Cuadras 1992)

Theorem 1 *Let (F, T, B) be a consistent 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*

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

where $d(i)$ denotes the degree of node i; i.e. 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 (F, T, t) be an n-dimensional rank correlation tree specification, then there exists a joint probability distribution G realizing (F, T, t) with G tree dependent.*

Theorem 2 would not hold if we replaced rank correlations with product moment correlations in Definition 4. Given 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.

Any random vector $\underline{X}$ with multivariate probability distribution function $F_{\underline{X}}$ can be obtained as a 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 the random variables $(Z_1, \ldots, Z_n, \mathcal{L})$ with a Markov tree dependent realization $G_{\underline{Z}, \mathcal{L}}$ such that $\int G_{\underline{Z}, \mathcal{L}}(\underline{x}, \ell)\, d\ell = F_{\underline{X}}(\underline{x})$.*

2. Entropy of Markov Tree Dependent Distributions

From Theorem 1 it follows immediately that for the Markov tree dependent density g given by the theorem:

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

If the bivariate tree specification does not completely specify the bivariate marginals $f_{i,j}$, $\{i,j\} \in E$, 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 $I(f_{i,j}|f_i f_j)$, $\{i,j\} \in E$.

In this section we show that Markov tree dependent distributions are optimal realizations of bivariate tree specifications in a maximal entropy sense (i.e. minimal relative information). In other words, we show that a maximal entropy realization of (consistent) bivariate tree specification has Markov tree dependence. This follows from a very general result stating that maximal entropy distributions, relative to independent distributions, subject to overlapping marginal constraints, are conditionally independent given the overlap.

To prove the theorem, we first formulate three lemmas. We assume in this analysis that the distributions have densities and that the absolute continuity condition is always fulfilled. 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 $g_X(x)$, $g_Y(y)$ and $g_Z(z)$; and bivariate marginals $g_{X,Y}$, $g_{X,Z}$. $g_{X|Y}$ and $g_{Y|X}$ are the conditional densities of X given Y and of Y given X respectively.

Lemma 4

Let

$$\tilde{g}_{X,Y,Z}(x,y,z) \begin{cases} = g_{X,Y}(x,y) g_{Z|X}(x,z) \\ = g_{X,Z}(x,z) g_{Y|X}(x,y) \ . \end{cases}$$

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 Y and Z are conditionally independent given X under $\tilde{g}$.

Proof

The proof is a straigthforward calculation. □

Lemma 5 *With g as above, let $p_X(x)$ be a density. Then*

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

and equality holds if and only if X and Y are independent under g; i.e. $g_{X|Y}(x,y) = g_X(x)$.

Proof

By definition

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

is equivalent to

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

$$\int g_X(x) \log \frac{g_X(x)}{p_X(x)} \ dx$$

or 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 as

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

This last equation equals $I(g_{X,Y}|g_X g_Y)$. It always holds and it holds with equality if and only if $g_{X,Y} = g_X g_Y$, (Kullback, 1959). This quantity is also called *mutual information.* □

Lemma 6

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 4, then

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

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

and 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 5 is used for the inequality. Hence

$$I(g_{X,Y,Z}|g_Xg_Yg_Z) \geq I(g_{XY}|g_Xg_Y) + I(g_{XZ}|g_Xg_Z) \quad (2)$$

with equality if and only if Z and Y are independent given X, which holds for $\tilde{g}$ (lemma 4). □

We may now formulate

Theorem 7 *Assume that $g_{X,Y}$ is a probability density with marginals f_X and f_Y that uniquely minimizes $I(g_{X,Y}|f_Xf_Y)$ 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 $I(g_{X,Z}|f_Xf_Z)$ within the class of distributions $B(X,Z)$. Then $g_{X,Y,Z} := g_{X,Y}g_{Z|X}$ is the unique probability density with marginals f_X, f_Y and f_Z that minimizes $I(g_{X,Y,Z}|f_Xf_Yf_Z)$ with the marginals $g_{X,Y}$ and $g_{X,Z}$ 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 $I(f_{X,Y,Z}|f_Xf_Yf_Z) \leq I(g_{X,Y,Z}|f_Xf_Yf_Z)$ Then by Lemma 4 we may assume without loss of generality that $f_{X,Y,Z} = \tilde{f}_{X,Y,Z} := f_{XY}f_{Z|X}$. By Lemma 6 we have

$$I(\tilde{f}_{X,Y,Z}|f_Xf_Yf_Z) = I(f_{X,Y}|f_Xf_Y) + I(f_{X,Z}|f_Xf_Z).$$

But

$$I(f_{X,Y}|f_Xf_Y) + I(f_{X,Z}|f_Xf_Z)$$

$$\geq I(g_{X,Y}|f_Xf_Y) + I(g_{X,Z}|f_Xf_Z) =$$

$$= I(g_{X,Y,Z}|f_Xf_Yf_Z) \geq$$

$$\geq I(\tilde{f}_{X,Y,Z}|f_Xf_Yf_Z) =$$

$$= I(f_{X,Y}|f_Xf_Y) + I(f_{X,Z}|f_Xf_Z)$$

.

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

Corollary 8 *Let (F,T,B) be a consistent bivariate tree specification. For each $(i,j) \in E$, let there be a unique density $g(x_i,x_j)$ which has maximal entropy relative to the product measure f_if_j under the constraint $B(ij)$. Then is the unique density with maximal entropy relative to the product density $\prod_{i\in N} f_i$ under constraints $B(i,j), \{i,j\} \in E$.*

Proof

Using the notation of Theorem 1, the proof is by induction on n. The densities g_j are the unique maximum entropy densities for the subtrees $T_j, j \in D_i$, by the induction hypothesis. If $g_{j|i} = g_j/f_i$, then the density $g = f_i\prod_{j\in D_i} g_{j|i}$, has maximal entropy by Theorem 7 under the constraints implied by T_j for all $j \in D_j$. These are the same constraints as (F,T,B). Hence, g is a maximal entropy realization of (F,T,B).

□

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 maximal entropy density given (F,T,B) and this density is Markov tree dependent.

3. 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 dimension. In addition to specifying marginals g_1, g_2, and g_3, and rank correlations $\rho_r(x_1, x_2), \rho_r(x_2, x_3)$, we also specify the conditional rank correlation as a function of x_2:

$$\rho_r(x_1, x_3|x_2) = \rho_r((x_1|x_2), (x_3|x_2)).$$

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 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 1 and 2 show (regular) vine specifications on 5 variables. Figure 3 shows an irregular vine specification. 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 top level restrictions on the bivariate marginals form a tree T_1 with nodes 1, ...5. The next level forms a three 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, $i = 1, ..n-1$ have maximal cardinality $n-i$, as we may "remove" any edge by associating with it the vacuous restriction.

The results obtained for vines are slightly different than the results of Theorem 7. There, we started with marginals satisfying the constraints and built a joint satisfying the tree specification. Now, we cannot rely on factoring the joint, and must settle for slightly weaker results: A joint satisfying the constraints is assumed; if the constrains are satisfied in a minimally informative way, then the joint is minimally informative. The following illustrates this in the notation of the previous section.

Theorem 9 *Let X, Y, Z be as in the previous section and let g be a density on X, Y, Z. Suppose that g_{YX} is the unique density satisfying B_{YX} and minimizing $I(g_{YX}|g_Y g_X)$, suppose that g_{ZX} is the unique density satisfying B_{ZX} and minimizing $I(g_{ZX}|g_Z g_X)$, and suppose that $g_{YZ|X}$ is the unique conditional density satisfying $B_{YZ|X}$ and minimizing $I(g_{Y|X}|g_{Y|X} g_{Z|X})$. Then g is the unique density satisfying these constraints and minimizing $I(g|g_X g_Y g_Z)$.*

[1] We ignore measurability constraints here

Proof
We sketch the proof, leaving details to the reader.

$$I(g|g_X g_Y g_Z) = I(g_{YX}|g_Y g_X) + E_{YX} I(g_{Z|YX}|g_Z) =$$
$$I(g_{ZX}|g_Z g_X) + E_{XZ} I(g_{Y|XZ}|g_Y),$$

where E_{ij} denotes expectation taken over i, j. Further

$$E_{YX} I(g_{Z|YX}|g_Z) + E_{XZ} I(g_{Y|X3}|g_Y) =$$
$$E_X I(g_{YZ|X}|g_{Y|X} g_{Z|X}) + I(g|g_X g_Y g_Z).$$

Adding the two expressions for $I(g|g_X g_Y g_Z)$,

$$I(g|g_X g_Y g_Z) = I(g_{YX}|g_Y g_X) + I(g_{XZ}|g_X g_Z) +$$
$$E_X I(g_{YZ|X}|g_{Y|X} g_{Z|X}).$$

The last term is the expectation of a non-negative function. By assumption, the terms on the right hand side above are uniquely minimized, under the constraints, by g.

□

Definition 7 (vine)
$V(n)$ is a vine on n elements if

1. *$V(n) = (T_1, ...T_{n-1})$*
2. *T_1 is a tree with nodes $N_1 = \{1, \ldots, n\}$ and edges E_1,*
3. *$T_i, i = \{2, \ldots, n-1\}$ is a tree with nodes $N_i = E_{i-1}$ and edges E_i, $\#E_i = n-i$, where $\#E_i$ is the cardinality of the set E_i.*

Definition 8 (constraint set)
$CV(n)$ is the constraint set associated with $V(n)$ if

1. *$CV(n) = \{C_{i,jk}, D_{i,jk}, A_{i,jk},; i = 1...n-1; \{j,k\} \in E_i\}$*
2. *$C_{1,\{jk\}} = \{j, k\}$. For $i = 2, ...n-1$; $C_{i,jk} = C_{i-1,j} \cup C_{i-1,k}$.*

   Note that $C_{i-1,j}$ is well defined, as $j \in E_{i-1}$ is an edge in T_{i-1}. When no confusion can arise, we write $C_{i,jk}$ instead of $C_{i,\{j,k\}}$. The same convention will apply for the following terms.

3. $D_{1,jk} = \oslash$, *for* $i = 2, ...n-1; D_{i,jk} = C_{i-1,j} \cap C_{i-1,k}$

4. $A_{1,jk} = \{j,k\}$; *for* $i = 1, ...n-1; A_{i,jk} = C_{i-1,j} \Delta C_{i-1,k}$, *where* Δ *denotes the symmetric difference.*

Definition 9 (vine specification)
$(F, \mathcal{V}, B)$ *is a vine specification if*

1. $F = (F_1, \ldots, F_n)$ *is a vector of distribution functions for random vectors* $(X_1, ...X_n)$ *with no common coordinates.*

2. $\mathcal{V}(n)$ *is a vine on n elements*

3. $B = \{B_{i,jk}; i = 1, ...n-1; \{j,k\} \in E_i\}$; *where for* $\{j,k\} \in E_1$ $B_{1,jk}$ *is a subset of the set of two dimensional marginals* F_{jk} *having marginals* F_j, F_k. *For* $i = 2, ...n-1$; *if* $A_{i,jk} \neq \oslash$ *and* $D_{i,jk} \neq \oslash$, *then* $B_{i,jk}$ *is a subset of the set of marginals on* $A_{i,jk}$ *conditional on* $D_{i,jk}$, *and* $B_{i,jk}$ *is vacuous otherwise, that is,* $B_{i,jk}$ *is the entire set of conditional marginals* $F_{A_{i,jk}|D_{i,jk}}$,

For the vines in Figures 1,2 and 3, and for edge $\{j,k\} \in E_i$, the term $A_{i,jk}$ appears to the left of the "|", and the term $D_{i,jk}$ appears to the right of the "|". $C_{i,jk}$ is the union all terms, left and right of "|".

We say that a distribution F realizes a vine specification, or exhibits vine dependence, if it satisfies all constraints. The notion of a vine specification is very general, and it is not always easy to identify the constraint sets visually. In practice it is convenient to specify large joint distributions in terms of (conditional) correlations, and this means that constraint sets will be such that $\#A_{i,jk} = 2$. Vines satisfying this condition are termed *regular* and a proximity property provides a mechanism for constructing regular vines.

Definition 10 (regularity and proximity)

1. $\mathcal{V}(n)$ *is regular if for* $i = 1, ...n-1, \{j,k\} \in E_i, \#A_{i,jk} = 2$

2. $\mathcal{V}(n)$ *has the proximity property if for* $i = 2, ...n-1$, *and* $\{a,b\} \in E_i; \#a\Delta b = 2$.

The proximity property states that two edges in T_{i-1} (nodes in T_i) are joined by an edge in T_i only if the edges in T_{i-1} share a common node in T_{i-2}. The proximity property does not restrict the tree T_1. The vine in Figure 3 does not satisfy the proximity property, as the edge labelled [1234|0] joins edges in T_2 which do not share a common node in T_2.

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 1 shows a regular vine that is uniquely determined, the regular vine in Figure 2 is not uniquely determined. The edge labelled [25|3] could be replaced by an edge [45|3].

Lemma 10 *If* $\mathcal{V}(n)$ *satisfies the proximity property, then it is regular. Moreover,* $\#C_{i,jk} = i+1$.

Proof
The proof is by induction on $i = 1, ...n-1$. The basis case is trivial; assume that $\#A_{m,jk} = 2$ for $m < i$. We claim that for $h \leq n-1$; $\#C_{h,jk} = h+1$. This is trivial for $h = 1$ and if two sets X, Y satisfy $\#X = \#Y = h$ and $\#X\Delta Y = 2$, then $\#X \cup Y = h+1$.

We now write

$$C_{i,ab} = C_{i-1,a_1a_2} \cup C_{i-1,b_1b_2}$$

$$= \overbrace{C_{i-2,a_{11}a_{12}}}^{T} \cup \overbrace{C_{i-2,a_{21}a_{22}}}^{U} \cup$$

$$\cup \overbrace{C_{i-2,b_{11}b_{12}}}^{V} \cup \overbrace{C_{i-2,b_{21}b_{22}}}^{W}$$

By induction $\#T\Delta U = 2$, and since the $\#T = \#U$, one element of their symmetric difference must be in T and one element must be in U. Similar reasoning holds for $V\Delta W$. By the proximity property, one of $\{T,U\}$ is equal to one of $\{V,W\}$. It follows that

$$\#(C_{i-2,a_{12}a_{12}} \cup C_{i-2,a_{21}a_{22}})\Delta$$

$$(C_{i-2,b_{12}b_{12}} \cup C_{i-2,b_{21}a_{22}}) = 2.$$

□

Recalling notation, g is a density on $\mathbb{R}^n$ for which all marginal and conditional marginal densities satisfy the absolute continuity conditions implicit in the relative information integrals. $g_{1,..k}$ denotes the marginal over $x_1, ...x_k$, $g_{1,..k-1|k,..n}$ denotes the marginal over $x_1, ...x_{k-1}$ conditional on $x_k, ...x_n$. $E_{1,...k}$ denotes expectation taken over $x_1, ...x_k$.

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

Lemma 11

1.

$$I(g|\prod_{i=1}^{n} g_i) = I(g_{k,...n}|\prod_{i=k}^{n} g_i)+$$

$$E_{k,...n}I(g_{1,...k-1|k,...n}|\prod_{i=1}^{k-1} g_i).$$

2.

$$I(g|\prod_{i=1}^{n} g_i) = \sum_{j=1}^{n-1} E_{1,...j}I(g_{j+1|1,...j}|g_{j+1}).$$

3.

$$E_{2,...n}I(g_{1|2...n}|g_1) + E_{1...n-1}I(g_{n|1...n-1}|g_n) =$$
$$= E_{2,...n-1}\left(I(g_{1,n|2,...n-1}|g_{1|2,...n-1}g_{n|2,...n-1})\right.$$
$$\left.+I(g_{1,n|2,...n-1}|g_1g_n)\right)$$

4.

$$2I(g|\prod_{i=1}^{n} g_i) = I(g_{2,...n}|\prod_{i=2}^{n} g_i)+$$

$$I(g_{1,...n-1}|\prod_{i=1}^{n-1} g_i)+$$

$$E_{2,...n-1}I(g_{1,n|2,...n-1}|g_{1|2,...n-1}g_{n|2,...n-1})+$$
$$I(g|g_1g_ng_{2,...n-1})$$

Proof
We indicate the main steps, leaving the computational details to the reader.

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

$$\frac{gg}{g_{2,...n}g_{1,...n-1}g_1g_n}.$$

This can be re-written as

$$\frac{g_{1,n|2,...n-1}}{g_{1|2,...n-1}g_{n|2,...n-1}}\frac{g_{1,n|2,...n-1}}{g_1g_n}.$$

Writing the log of is product as the sum of logarithms of its terms, the result on the right hand side is obtained.

4. This follows from the previous statement by noting

$$E_{2,...n-1}I(g_{1,n|2,...n-1}g_1g_n) = I(g|g_1g_2g_{3,...n-1}).$$

□

Theorem 12 *Let g be an $n-$dimensional density satisfying the regular vine specification $(F, \mathcal{V}, B)$, and suppose $g_{A_{i,jk}|D_{i,jk}}$ is the unique density which minimizes*

$$I(g_{A_{i,jk}|D_{i,jk}}|g_{A1_{i,jk}|D_{i,jk}}g_{A2_{i,jk}|D_{i,jk}});$$
$$i = 1, ...n-1; \{jk\} \in E_i;$$

where $Ai, jk = \{A1_{i,jk}, A2_{(i,jk}\}$. Then g is the unique density satisfying $(F, \mathcal{V}, B)$ and minimizing

$$I(g|\prod_{i=1}^{n} g_i).$$

Proof
The proof is by induction on n. E_{n-1} has one element, say $\{1,2\}$, and we may assume that $A1_{n-1,12} = x_1$, $A2_{n-1,12} = x_n$. We define a vine specification $(F^1, \mathcal{V}^1, B^1)$ on $\{x_2, ...x_n\}$:

$$\begin{aligned} F^1 &= F_2, ...F_n; \\ N_i^1 &= N_i^1 \backslash x_1; \\ E_i^1 &= E_i\backslash\{j,k\} \text{ if } j = x_1 \text{ or } k = x_1; \\ B_{i,jk}^1 &= B_{i,jk} \text{ if } C_{i,jk} \subset \{x_2, ...x_n\}. \end{aligned}$$

We define vine specifications $(F^n, \mathcal{V}^n, B^n)$ on $\{x_1, ...x_{n-1}\}$ and $(F^{1,n}, \mathcal{V}^{1,n}, B^{1,n})$ on $\{x_2, ...x_{n-1}\}$ in the same manner. From the definition of regularity it follows immediately that $\mathcal{V}^1$, $\mathcal{V}^n$ and $\mathcal{V}^{1,n}$ are regular. $g_{2,...n}$, $g_{1,...n-1}$ and $g_{2,...n-1}$ satisfy the conditions of the theorem for these specifications. In other words,

$$I(g_{1,...n-1}|\prod_{i=1}^{n-1} g_i)$$

is minimal for densities satisfying B^n, and

$$I(g_{2,...n}|\prod_{i=2}^{n} g_i)$$

is minimal for densities satisfying B^1.
We now claim that

$$I(g|\prod_{i=1}^{n} g_i) = \sum_{\substack{i=1,...n-1 \\ \{j,k\}\in E_i}} E_{D_{i,jk}}$$

$$I(g_{A_{i,jk}|D_{i,jk}}|g_{A1_{i,jk}|D_{i,jk}}g_{A2_{i,jk}|D_{i,jk}}).$$

The claim is proved by applying lemma 11(4); the last term in the above sum is the expectation in lemma 11(4). Applying the induction hypothesis to the vine specification $(F^{1,n}, \mathcal{V}^{1,n}, B^{1,n})$ we note that the terms in the expansion of $I(g_{2,\dots n-1}|\prod_{i=2}^{n-1} g_i)$ are exactly those terms which are counted twice in the expansion of

$$I(g_{2,\dots n}|\prod_{i=2}^{n} g_i) + I(g_{1,\dots n-1}|\prod_{i=1}^{n-1} g_i).$$

from which the claim follows.

Since g minimizes each information term in equation(3.), it also minimizes each expectation and the theorem is proved. □

We note that if any constraint $B_{i,jk}$ in a regular vine is vacuous, then the variables $A_{i,jk}$ are conditionally independent given $D_{i,jk}$. Hence, if we represent a joint distribution via a regular vine, and indicate which constraints are vacuous, then we can immediately infer the corresponding conditional independence statements.

4. Acknowledgement

The author would like to thank Tim Bedford for reading the proofs and suggesting many improvements and extensions. Discussions with Hans Meeuwissen are also gratefully acknowledged

References

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

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

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

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

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

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

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

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

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

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

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

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

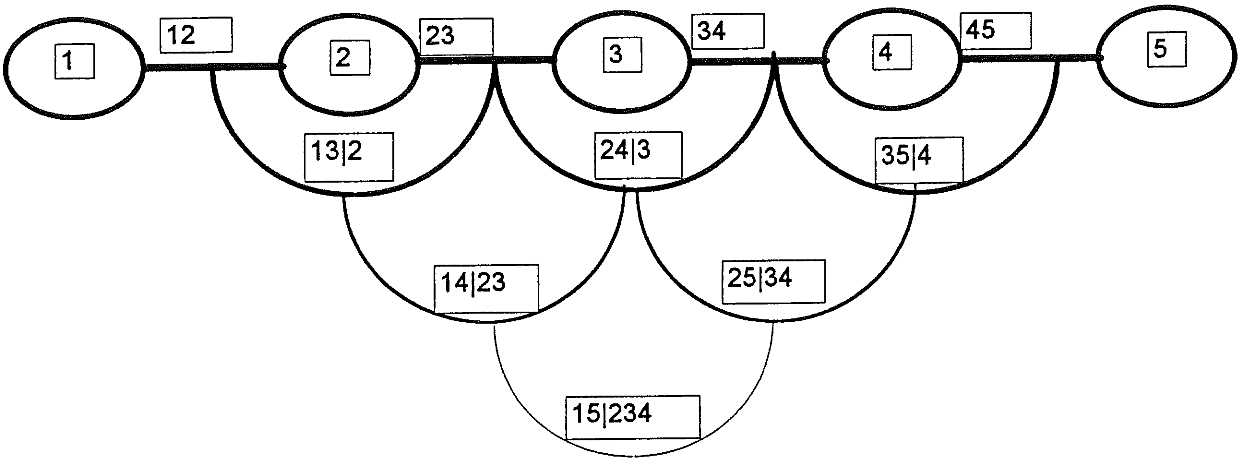

Figure 1
A (unique) regular vine on 5 elements

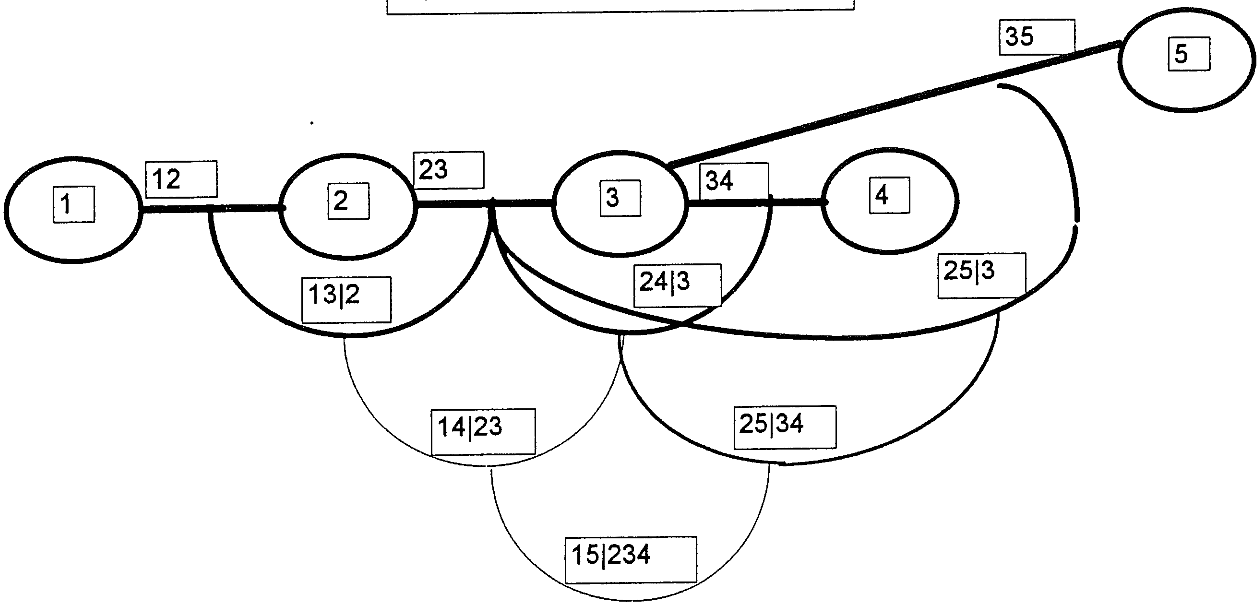

Figure 2
A (non-unique) regular vine on 5 elements

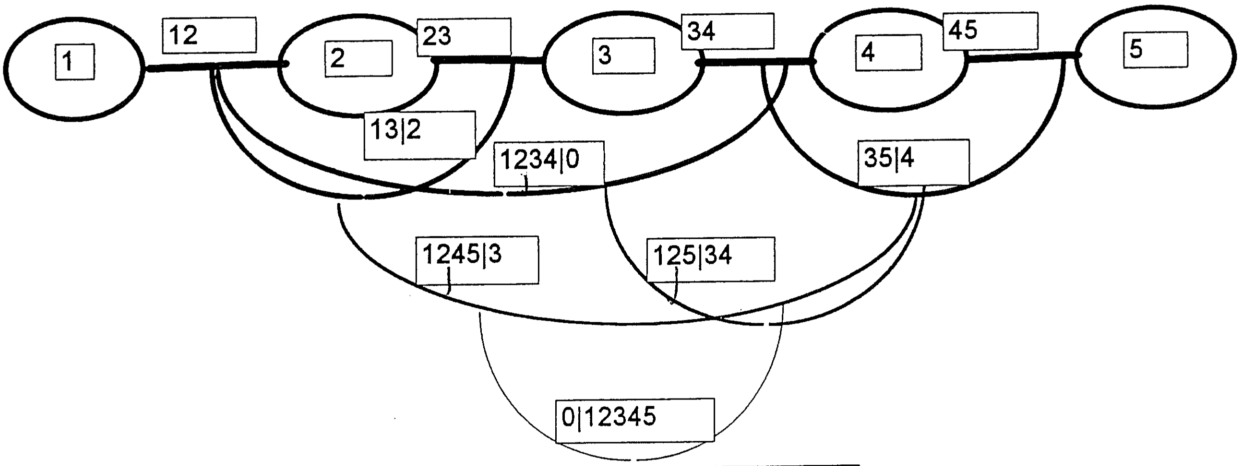

Figure 3
An irregular vine on 5 elements