DELFT UNIVERSITY OF TECHNOLOGY

REPORT 94-28

Tree dependent random variables

- revised version -

Adrianus M.H. Meeuwissen
Roger M. Cooke

ISSN 0922-5641

Reports of the Faculty of Technical Mathematics and Informatics no. 94-28

Delft 1994

Copyright © 1994 by Faculty of Technical Mathematics and Informatics, Delft, The Netherlands.
No part of this Journal may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from Faculty of Technical Mathematics and Informatics, Delft Univeersityof Technology, The Netherlands.

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

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

1 Introduction

We are interested in Monte Carlo simulation of high dimensional joint distributions in which dependencies may be important. Most current simulation programs generate correlated samples by transforming to a Gaussian distribution (Iman et al. (1981)), or by a "distribution-free" transformation of the sample (Iman and Conover (1982)). We do not assume that the joint distribution is a transform of the joint normal. Because of the high dimensionality in combination with a lack of data and/or a lack of knowledge a complete charaterization of the joint distributions is often not available or very hard to give. Even specifying the entire covariance matrix may be impracticable. We are therefore interested in convenient methods for partially specifying a joint distribution, where "convenient" means both convenient for the analyst modelling a given problem, and convenient for the computer in performing Monte Carlo simulation. A partial specification fixes certain properties. Most important for the present study is the rank correlation tree specification for n variables. It specifies:

1. marginal distributions for each of the n variables,
2. a tree of dependence relations i.e. an undirected acyclic graph on n points,
3. values in $[-1, 1]$ for each of the edges in the tree, such a value specifies the rank correlation coefficient between the variables connected via the edge.

A rank correlation tree is shown in Figure 1. More generally a bivariate tree specification associates arbitrary restrictions on the bivariate distribution of two connected variables.

We say that a distribution on $\mathbb{R}^n$ satisfies or realizes a bivariate tree specification if it has one-dimensional marginal distributions and bivariate marginal distributions that agree with the specifications. A bivariate tree specification is called consistent if it is realized by at least one multivariate probability distribution. If a bivariate tree specification is consistent (often

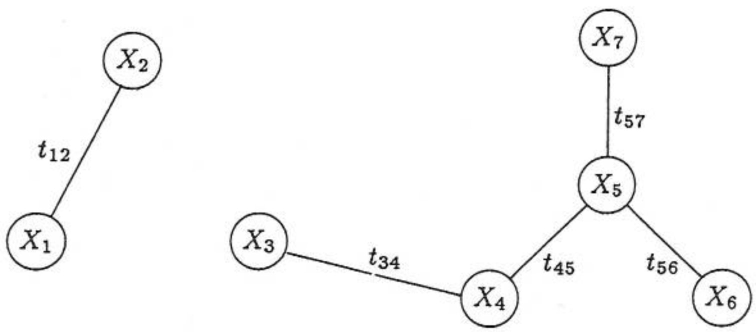

Figure 1: A rank correlation tree specification with 7 variables. Specified are all the one dimensional marginal distributions $F_1, \ldots, F_7$ and the rank correlations t for the variables that are neighbours, $t \in \{t_{12}, t_{34}, t_{45}, t_{56}, t_{57}\}$.

a non-trivial question) then generally, there will be many different multivariate distributions that satisfy the specifications. We show in this paper that the distribution having maximal entropy within the class of all distributions that realize a given consistent bivariate tree specification has many desirable properties. We make use of the following Theorem proven in Meeuwissen (1993).

Theorem 1 *Given two invertible, continuous one-dimensional marginal distributions F_1 and F_2 and a set of feasible moment restrictions on the joint distribution, there exists a unique bivariate distribution F having maximum entropy in the class of joint distributions with the prescribed one-dimensional marginals F_1 and F_2 and satisfying the restrictions.*

In this context, maximal entropy must be interpreted as minimal relative information with respect to the product distribution of F_1 and F_2. A similar theorem with only a finite number of moment restrictions and no fixed marginal distributions was already proven by Kullback (1959). Note that the specification of marginal distributions is not equivalent with specifying a finite number of moments. Examples of moment restrictions are e.g. cross-product moments, quantile constraints, rank correlation and expectations of functions of marginals. Further it was shown in Meeuwissen (1993) that bivariate maximum entropy distributions satisfying marginal distributions and a (rank) correlation constraint can be conveniently approximated and simulated. We summarize some results of this paper informally. In this paper we assume that all distributions have densities.

1. A rank correlation tree specification is always consistent.
2. Every consistent bivariate tree specification has a realization that factors into bivariate and univariate distributions; i.e. it has a realization with the property that two sets of variables on one path are independent conditional on any variable on the path in the tree which separates these sets. Two variables not connected by a path are independent. Such distributions are called Markov tree dependent.
3. Given a consistent bivariate tree specification for n variables, where Theorem 1 is applicable to the partially specified bivariate distributions, there is a unique joint distribution

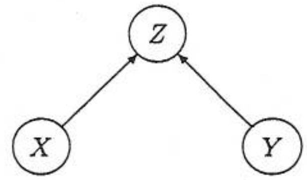

Figure 2: An influence diagram of three variables; X and Y "influence" Z.

having maximal entropy in the class of distributions realizing this bivariate tree specification. Moreover this distribution has the Markov tree dependence property mentioned in 2. Its bivariate distributions are determined by the unique bivariate maximal entropy distributions.

4. Given a completely characterized distribution F on n variables, there is a consistent bivariate tree specification on $n+1$ variables such that variables $1, \ldots, n$ are conditionally independent given variable $n+1$, and the n-dimensional marginal distribution of the first n variables equals F.

Result 3 implies that finding the "least informative" multivariate probability distribution complying with the bivariate tree specification can be simplified to the calculation of least informative bivariate distributions.

The Markov tree structure of the partial specification describes a sampling strategy in the obvious way: Pick a "root" arbitrarily, hang the tree from its root, and sample each variable conditional on the variable hanging above it. The whole distribution is simulated by conditionalizing two dimensional distributions.

The trees appearing in rank correlation tree specifications and bivariate tree specifications are closely related to influence diagrams, Oliver & Smith (1990). Influence diagrams however, are described by directed acyclic graphs, where the direction of an arc in the graph is taken to describe the flow of "influence" (for a discussion of the differences between directed and undirected graphs in specifying a probability model, see Speed, 1990). In the present context, the reason for studying undirected rather than directed graphs may be explained with the help of the influence diagram in Figure 2. According to the prevailing interpretation Figure 2 says that X and Y are independent. To simulate the distribution $F_{X,Y,Z}$, Z has to be sampled conditional on both X and Y; i.e. this distribution does not factor into two- and one dimensional distributions. More generally, an influence diagram does not guarantee a lower dimensional factorization of the joint distribution. For purposes of simulation, it is precisely this property which we wish to have.

The paper is structured as follows. In Section 2 the definitions and preliminaries of the concept of bivariate tree specifications and tree dependence are introduced. Then in Section 3 we prove results 1, 2 and 4. In Section 4 the entropy result 3 is proven. In Section 5 total positivity and regression dependence properties of the maximal entropy multivariate distributions are studied. Section 6 concludes the paper with some remarks on possible generalizations.

2 Definitions and Preliminaries

When we speak of random variables or random vectors, a probability space will always be assumed but seldom specified. Usually we consider continuous invertible probability distributions F on $\mathbb{R}^n$ equipped with the Borel sigma algebra $\mathcal{B}$. The marginal distribution functions of F are denoted with F_i $(1 \leq i \leq n)$ and the bivariate marginal distribution functions with F_{ij} $(1 \leq i \neq j \leq n)$.

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 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 of two random variables is the ordinary product moment correlation of two transformations of these random variables. It is sometimes called grade correlation. An extensive discussion can be found in Kruskal (1958). The probability distribution functions F_X and F_Y transform X and Y respectively to uniform(0,1) variables U and V (if F_X and F_Y are continuous). Calculation of the integral

$$E\{F_X(X)F_Y(Y)\} = \int\int F_X(x)F_Y(y)f_{X,Y}(x, y)dxdy$$

is the basic step in the calculation of $\rho_r(X, Y)$ as we know that a uniform(0,1) variable U has mean $1/2$ and variance $1/12$. 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 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 (rank correlation tree specification)
(F, T, t) is an n-dimensional rank correlation tree specification if:

1. $F = (F_1, \ldots, F_n)$ *is a vector of completely characterized one-dimensional distribution functions,*

2. T *is an acyclic undirected graph with nodes* $N = \{1, \ldots, n\}$ *and edges* E, *where an edge is an unordered pair* $\{i, j\}$ *with* i *and* $j \in N$,

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 4 (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 completely characterized one-dimensional distribution functions,*

2. T *is an acyclic undirected graph with nodes* $N = \{1, \ldots, n\}$ *and edges* E *where an edge is an unordered pair* $\{i, j\}$ *with* i *and* $j \in N$,

3. B *restricts the bivariate distributions* F_{ij}, $\{i, j\} \in E$ *to belong to a non-empty subclass* $F_{B(ij)}$ *of the class of distribution functions with marginals* F_i *and* F_j.

Definition 5 (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 $(1 \leq i \leq n)$ *and if for* $\{i, j\} \in E$ *the bivariate distributions* G_{ij} *of* G *are members of the subclass* $F_{B(ij)}$.
(ii) G *has tree dependence described by* T *if* $\{i, k_1\}, \ldots, \{k_m, j\} \in E$ *implies that* X_i *and* X_j *are conditionally indepedent given* k_ℓ, *for each* ℓ, *taken singly,* $1 \leq \ell \leq m$; *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 described by* T *if for all* $i \in N$, *the following property holds: Let* J *and* K *be disjoint subsets of* N *not containing* i *such that for all* $j \in J, k \in K$, j *and* k *are connected by a path containing* i *with* i *between* j *and* k; *and let* $X_J = \{X_l : l \in J\}, X_K = \{X_l : l \in K\}$; *then* X_J *and* X_K *are independent conditional on* X_i.

Markov tree dependence entails tree dependence, but not conversely; the following example shows that tree dependence is not equivalent with Markov tree dependence. Let X_1, X_2, X_3 be independent Bernoulli variables, taking values in $\{0, 1\}$, and Let $X_4 = 1$ if $X_1 + X_2 + X_3$ is even and $=$ 0 otherwise. Then X_1 and X_4 are independent given X_2, and are independent given X_3, but X_1 and $\{X_3, X_4\}$ are not independent given X_2. For trees in which all paths have length at most three, tree dependence and Markov tree dependence coincide.

Note that tree dependence says nothing about marginals or correlations; rather it describes a conditional independence structure. Satisfying a rank correlation tree specification does not imply tree dependence. Indeed, every distribution on $\mathbb{R}^n$ satisfies rank correlation tree specifications, but not every distribution exhibits conditional independence relations. It suffices to consider three variables, no two of which are independent conditional on the third. Finally, we remark that the existence of an edge $\{i, j\} \in E$ does not imply that variables X_i and X_j are not independent. Indeed, if $t_{ij} = 0$, then the maximal entropy realization of (F, T, t) will make X_i and X_j independent.

3 Existence of Markov Tree Dependent Distributions

In this section we prove some basic properties of Markov tree dependent random variables. Theorem 2 is similar to results known from the study of influence diagrams. Theorem 3 shows that correlation tree specifications have Markov tree dependent realizations. Further Theorem 4 shows that any n-dimensional joint probability distribution can be constructed as an n-dimensional marginal distribution from a Markov tree dependent $(n+1)$-dimensional distribution.

Theorem 2 *Let (F,T,B) be a consistent n-dimensional bivariate tree specification that specifies the marginal densities f_i, $1 \leq i \leq n$ 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*

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

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

Proof
The proof is by induction on n. If $n = 1$ the theorem is trivial. Without loss of generality assume that T is connected. Fix $i \in N$ with degree at least 2 and let D_i denote the set of neighbours of i: $D_i = \{j | j \in N, \{i,j\} \in E\}$. Now consider for each $j \in D_i$ the subtrees T_j with set of nodes $N_j = \{i,j\} \cup \{k | k \in N$, there is a path from k to j and this path does not include $i\}$ and set of edges $E_j = \{\{k,\ell\} | \{k,\ell\} \in E,\ k,\ell \in N_j\}$. Thus $\cup_{j\in D_i} E_j = E$ and for all $j,k \in D_i$: $E_j \cap E_k = \emptyset$ and $N_j \cap N_k = \{i\}$ because T is a tree. Further let $d_j(k)$ denote the degree of node k in T_j; i.e. $d_j(i) = 1$ and $d_j(k) = d(k)$ for all other $k \in N_j$. Now let g_j be the unique distribution satisfying the theorem for the subtree T_j for all $j \in D_i$. By induction we have for $N_j = \{\ell_1, \ell_2, \ldots, \ell_{n(j)}\}$

$$g_j(x_{\ell_1}, x_{\ell_2}, \ldots, x_{\ell_{n(j)}}) = \frac{\prod_{(\ell_i,\ell_j)\in E_j} f_{\ell_i,\ell_j}(x_{\ell_i}, x_{\ell_j})}{\prod_{h\in N_j}(f_{\ell_h}(x_{\ell_h}))^{d_j(\ell_h)-1}}\,.$$

This expression is equivalent to

$$g_j(x_{\ell_1}, x_{\ell_2}, \ldots, x_{\ell_{n(j)}}) = f_{i,j}(x_i, x_j)\frac{\prod_{(\ell_i,\ell_j)\in E_j\setminus\{\{i,j\}\}} f_{\ell_i,\ell_j}(x_{\ell_i}, x_{\ell_j})}{\prod_{h\in N_j\setminus\{i\}}(f_{\ell_h}(x_{\ell_h}))^{d(\ell_h)-1}}\,.$$

Denote the conditional density of g_j given X_i as $g_{j|i}$; i.e. $g_{j|i} = g_j/f_i$. Since the sets $N_j \setminus \{i\}$, $j \in D_i$ are disjunct we may write

$$g(x_1, \ldots, x_n) = f_i(x_i) \prod_{j\in D_i} g_{j|i}(x_{\ell_1}, \ldots, x_{\ell_{n(j)}})$$

which equals (1). To verify that g has the Markov tree dependence property, note that the above equaton follows from (1) for any $i \in N$ of degree at least two. Let i, J, K be as in the definition of Markov tree dependence. For $j \in T_j, k \in T_j$, there is a path from k to j through j which does not contain i. Therefore

$$J \cap T_j \neq \emptyset \Longrightarrow K \cap T_j = \emptyset.$$

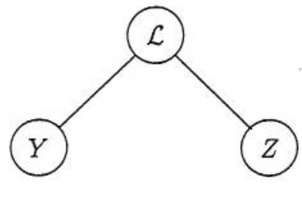

Figure 3: A bivariate tree specification on $(Y, Z, \mathcal{L})$.

Put $J_j = J \cap T_j, K_j = K \cap T_j; j \in D_i$; then

$$g(x_J, x_K|x_i) = \prod_{j:J_j \neq \emptyset} g_{j|i}(x_{J_j}) \prod_{k:K_j \neq 0} g_{k|i}(x_{K_k})) = g(x_J|x_i)g(x_K|x_i)).$$

To verify that g is unique, let $\tilde{g}$ be another density satisfying the theorem, let x_1 be a node with degree 1, and let $\{1,2\}$ be the edge attached to x_1. Let $I = N \setminus \{1,2\}$. Then

$$g(x_1, \ldots x_n) = g(x_I|x_2)g(x_2)g(x_1|x_2),$$

with a similar equation for $\tilde{g}$. By the induction hypothese, and by the equality of the first and second dimensional marginals, it follows that

$$g(x_1, \ldots x_n) = \tilde{g}(x_1, \ldots x_n)$$

□

Remark, Because of the general applicability of the induction argument of the proof of Theorem 2 it is in many cases sufficient to consider bivariate tree specifications of only three variables $(Y, Z, \mathcal{L})$, see Figure 3. The following theorem states that a rank correlation tree specification is always consistent.

Theorem 3 *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.*

Proof

A standard construction using Fréchet bounds, see e.g. Cuadras (1992), realizes bivariate distributions with uniform(0,1) marginals and any (rank) correlation value $t \in [-1, 1]$. Transformation of the uniform distributions to the required marginals F_i preserves the rank correlations t $(1 \leq i \leq n)$. Now consistent bivariate distributions F_{ij} with prescribed marginals and rank correlations are constructed, we can apply Theorem 2.

□

Theorem 3 would not hold if we replaced rank correlations with product moment correlations in Definition 3. 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.

Now we show that 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 $(\underline{X}, \mathcal{L})$.The extra variable $\mathcal{L}$ is also called a latent variable.

Theorem 4 *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 $\int G_{\underline{Z},\mathcal{L}}(\underline{x}, \ell)\ d\ell = F_{\underline{X}}(\underline{x})$.*

Proof
Let $M : \mathbb{R}^n \to \mathbb{R}$ be a bijective measurable map from $\mathbb{R}^n$ to $\mathbb{R}$. For the existence of such a map we refer to any standard textbook of measure theory, (eg Halmos (1950) p. 153). Define $\mathcal{L} = M(\underline{X})$. Now conditional upon $\mathcal{L}$ all X_i are deterministically known and hence they are in trivially independent. All paths in this tree have length three, hence the tree dependence is Markov. □

4 Entropy of Markov Tree Dependent Distributions

From Theorem 2 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.

Lemma 5
Let $g_{X,Y,Z}(x,y,z)$ be a density with marginal densities $g_X(x)$, $g_Y(y)$ and $g_Z(z)$; and 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 6 *Let $g_{X,Y}(x,y)$ be a probability density with marginals $g_X(x)$ and $g_Y(y)$ respectively, and let $p_X(x)$ be a density. Let $g_{X|Y}$ and $g_{Y|X}$ denote the conditional densities of X given Y and of Y given X respectively. 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 7

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 5, then

$$\begin{array}{ll} i) & I(g_{X,Y,Z}|g_X g_Y g_Z) \geq I(\tilde{g}_{X,Y,Z}|g_X g_Y g_Z)\ , \\ ii) & I(\tilde{g}_{X,Y,Z}|g_X g_Y g_Z) = I(g_{X,Y}|g_X g_Y) + I(g_{X,Z}|g_X g_Z)\ . \end{array}$$

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

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

and with

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

Now rewrite this as follows

$$I(g_{X,Y}|g_X g_Y) + \int g_X(x) \int g_{Y|X}(x,y) I(g_{Z|X,Y}|g_Z) \, dydx \; ,$$

and apply Lemma 6 to obtain for this expression the lower bound

$$I(g_{X,Y}|g_X g_Y) + \int g_X(x) I(g_{Z|X}|g_Z) \, dx$$

which can be rewritten as

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

or as

$$I(g_{X,Y}|g_X g_Y) + I(g_{X,Z}|g_X g_Z) \; .$$

This lower bound is obtained if and only if g is such that Z and Y are independent given X under g (lemma 6). This is indeed the case for $\tilde{g}$ by lemma 5. □

We may now formulate

Theorem 8 *Let $g_{X,Y}$ be the unique probability density with marginals f_X and f_Y that minimizes $I(g_{X,Y}|f_X f_Y)$ within the class of distributions $F_{B(X,Y)}$. Let similarly $g_{X,Z}$ be the unique probability density with marginals f_X and f_Z that minimizes $I(g_{X,Z}|f_X f_Z)$ within the class of distributions $F_{B(X,Z)}$. 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 $I(g_{X,Y,Z}|f_X f_Y f_Z)$ with the marginals $g_{X,Y}$ and $g_{X,Z}$ members of $F_{B(X,Y)}$ and $F_{B(X,Z)}$ respectively.*

Proof
Let $f_{X,Y,Z}$ be the joint probability density with marginals f_X, f_Y and f_Z that minimizes $I(f_{X,Y,Z}|f_X f_Y f_Z)$ given the constraints $B(X,Y)$ and $B(X,Z)$. Then by Lemma 5 $\tilde{f}_{X,Y,Z} := f_{XY} f_{Z|X}$ satisfies the same constraints B By Lemma 7 we have

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

But by the minimality of $f_{X,Y,Z}$ we also have .

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

Hence by Lemma 7 $f_{X,Y,Z} = \tilde{f}_{X,Y,Z}$, and

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

But

$$I(f_{X,Y}|f_X f_Y) + I(f_{X,Z}|f_X f_Z) \geq I(g_{X,Y}|f_X f_Y) + I(g_{X,Z}|f_X f_Z) = I(g_{X,Y,Z}|f_X f_Y f_Z)$$

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

Corollary 9 *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 $f_i f_j$ under the constraint $B(ij)$. Then (1) 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 2, 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 8 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.

5 Tree Dependent Random Variables and Total Positivity

The following theorems describe the 'smoothness' of tree dependent multivariate distributions. This 'smoothness' is related to the maximal entropy property of these distributions.

Definition 6 (total positivity of degree n)
Let $n \in \{2, 3, 4, \ldots\}$. A density $f_{X,Y}(x, y)$ is called totally positive of degree n (TPn) if for all $x_1 \leq x_2 \leq \ldots \leq x_n$ and for all $y_1 \leq y_2 \leq \ldots \leq y_n$, the matrix $\mathbf{M}$ *with elements $m_{i,j} = f_{X,Y}(x_i, y_j)$ satisfies*

$$det(\mathbf{M}) \geq 0 \ . \qquad (2)$$

If the same equation holds with the $\geq$ sign reversed, $f_{X,Y}$ is called totally negative of degree n (TNn), see Karlin (1968) and Marshall & Olkin (1979).

Total positivity of degree n of a distribution may be associated with 'smoothness' of the distribution. See Hutchinson & Lai (1990) for a brief overview of concepts of smoothness. Total positivity is an important, very strong property of joint probability distributions. Most of the concepts of dependence are implied by total positivity of degree 2. Note that if X, Y are independent, then $f_{X,Y}$ is TP2. Also the joint Gaussian distribution with positive correlation coefficient is TP2 as shown by Tong (1990).

Theorem 10 *Let $\mathcal{L}$ be a random variable with compact support and strictly positive density $f_{\mathcal{L}}(x)$. If both the joint distribution of the random variables $\mathcal{L}$ and Y and the joint distribution of the random variables $\mathcal{L}$ and Z are TP2, and Y and Z are independent conditional on $\mathcal{L}$, then the joint distribution of the random vector (Y, Z) is TP2.*

Proof. Assume $y_1 < y_2,\ z_1 < z_2$ and let $f_{\mathcal{L},Y}(\ell, y)$ and $f_{\mathcal{L},Z}(\ell, z)$ be TP2. We show:

$$f_{Y,Z}(y_1, z_1) f_{Y,Z}(y_2, z_2) - f_{Y,Z}(y_1, z_2) f_{Y,Z}(y_2, z_1) \geq 0 \ .$$

Conditionalize on $\mathcal{L}$ to get the integral

$$\int\!\!\int \Big[f_{Y,Z|\mathcal{L}}(\ell_1, y_1, z_1) f_{Y,Z|\mathcal{L}}(\ell_2, y_2, z_2) - f_{Y,Z|\mathcal{L}}(\ell_1, y_1, z_2) f_{Y,Z|\mathcal{L}}(\ell_2, y_2, z_1) \Big] f_{\mathcal{L}}(\ell_1) f_{\mathcal{L}}(\ell_2)\ d\ell_1 d\ell_2 \ .$$

In this expression one can exchange the names of the integration variables ℓ_1 and ℓ_2. Thus obtain the symmetric expression

$$\frac{1}{2} \int\!\!\int \Big\{ \Big[f_{Y,Z|\mathcal{L}}(\ell_1, y_1, z_1) f_{Y,Z|\mathcal{L}}(\ell_2, y_2, z_2) - f_{Y,Z|\mathcal{L}}(\ell_1, y_1, z_2) f_{Y,Z|\mathcal{L}}(\ell_2, y_2, z_1) \Big] + \Big[f_{Y,Z|\mathcal{L}}(\ell_2, y_1, z_1) f_{Y,Z|\mathcal{L}}(\ell_1, y_2, z_2) - f_{Y,Z|\mathcal{L}}(\ell_2, y_1, z_2) f_{Y,Z|\mathcal{L}}(\ell_1, y_2, z_1) \Big] \Big\} f_{\mathcal{L}}(\ell_1) f_{\mathcal{L}}(\ell_2)\ d\ell_1 d\ell_2 \ .$$

Use the conditional independence on $\mathcal{L}$ to rewrite this as

$$\frac{1}{2} \int\!\!\int \Big\{ \Big[f_{Y|\mathcal{L}}(\ell_1, y_1) f_{Z|\mathcal{L}}(\ell_1, z_1) f_{Y|\mathcal{L}}(\ell_2, y_2) f_{Z|\mathcal{L}}(\ell_2, z_2) - f_{Y|\mathcal{L}}(\ell_1, y_1) f_{Z|\mathcal{L}}(\ell_1, z_2) f_{Y|\mathcal{L}}(\ell_2, y_2) f_{Z|\mathcal{L}}(\ell_2, z_1) \Big] + \Big[f_{Y|\mathcal{L}}(\ell_2, y_1) f_{Z|\mathcal{L}}(\ell_2, z_1) f_{Y|\mathcal{L}}(\ell_1, y_2) f_{Z|\mathcal{L}}(\ell_1, z_2) - f_{Y|\mathcal{L}}(\ell_2, y_1) f_{Z|\mathcal{L}}(\ell_2, z_2) f_{Y|\mathcal{L}}(\ell_1, y_2) f_{Z|\mathcal{L}}(\ell_1, z_1) \Big] \Big\} f_{\mathcal{L}}(\ell_1) f_{\mathcal{L}}(\ell_2)\ d\ell_1 d\ell_2 \ .$$

Now write $f_{Y|\mathcal{L}}(\ell, y) = f_{\mathcal{L},Y}(\ell, y)/f_{\mathcal{L}}(\ell)$ and write the expression between braces as a product in the following way

$$\frac{1}{2} \int\!\!\int [f_{\mathcal{L},Y}(\ell_1, y_1) f_{\mathcal{L},Y}(\ell_2, y_2) - f_{\mathcal{L},Y}(\ell_2, y_1) f_{\mathcal{L},Y}(\ell_1, y_2)] [f_{\mathcal{L},Z}(\ell_1, z_1) f_{\mathcal{L},Z}(\ell_2, z_2) - f_{\mathcal{L},Z}(\ell_1, z_2) f_{\mathcal{L},Z}(\ell_2, z_1)] \frac{1}{f_{\mathcal{L}}(\ell_1) f_{\mathcal{L}}(\ell_2)}\ d\ell_1 d\ell_2 \ .$$

Now by assumption the integrand is a product of nonnegative factors for $\ell_1 < \ell_2$; and if $\ell_1 > \ell_2$ both factors between brackets change sign. Hence the expression is nonnegative. $\square$

Note: A similar, somewhat less general theorem can be found in Holland & Rosenbaum (1986).

Regression dependence is one of the concepts of dependence that is implied by total positivity of degree 2. Various definitions of regression dependence exist. They are not all entirely equivalent. We use here the following definition which is slightly weaker than other definitions, Lehmann (1963).

Definition 7 (Regression Dependence)
A random variable Y is positively (negatively) *regression dependent with respect to a random variable X if the conditional expectation $E\{Y|X = x\}$, also called the regression of Y with respect to X, is non-decreasing (non-increasing) in x.*

If the joint probability density of $f_{X,Y}$ is TP2, then is Y positively regression dependent with respect to X and is X positively regression dependent with respect to Y, Hutchinson & Lai (1990). When considering models that are monotone increasing/decreasing in the variables the following properties can be proven.

Theorem 11 *Let each of the components X_i, $1 \leq i \leq n$, of the stochastic vector $\underline{X} \in \mathbb{R}^n$ be positively regression dependent with respect to a stochastic variable $\mathcal{L}$ and let each pair (X_i, X_j) be independent conditional on $\mathcal{L}$, $1 \leq i, j \leq n$, $i \neq j$. If the function $M : \mathbb{R}^n \to \mathbb{R}$ is* non-decreasing *in each of its arguments then is $Y = M(\underline{X})$ positively regression dependent with respect to $\mathcal{L}$.*

Proof. We show that

$$E\{Y|\mathcal{L} = \ell\}$$

is non-decreasing in ℓ, if each of the random variables X_i is positively regression dependent with respect to $\mathcal{L}$ and M is non-decreasing.

$$E\{Y|\mathcal{L} = \ell\} = E\{M(X_1, \ldots, X_n)|\mathcal{L} = \ell\}$$

which is equivalent to

$$\int \cdots \int M(x_1, \ldots, x_n) f_{X_1,\ldots,X_n|\mathcal{L}}(x_1, \ldots, x_n|\ell)\, dx_1 \cdots dx_n \,.$$

Because of the conditional independence (on $\mathcal{L}$) of $X_1, \ldots, X_n$, we can rewrite this as

$$\int \cdots \int M(x_1, \ldots, x_n) f_{X_1|\mathcal{L}}(x_1|\ell) \cdots f_{X_n|\mathcal{L}}(x_n|\ell)\, dx_1 \cdots dx_n$$

or as

$$\int \cdots \left\{ \int M(x_1, \ldots, x_n) f_{X_1|\mathcal{L}}(x_1|\ell) dx_1 \right\} \cdots f_{X_n|\mathcal{L}}(x_n|\ell) dx_n \,,$$

and M is non-decreasing in all of its arguments. $\square$

6 Concluding Remarks

The introduced bivariate tree specifications are a formalism to handle the situation of "partially known" joint probability distributions that arise in Monte Carlo simulation studies. The fact that these specifications have Markov tree dependent realizations gives a direct sampling scheme. Moreover we have shown that these Markov tree dependent realizations have a maximal entropy property 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) van Dorp (1991) and Cooke (1995).

We have also given a theoretical construction that shows that any n-dimensional joint probability distribution can be obtained as a marginal distribution of a tree dependent realization of an $(n+1)$-dimensional bivariate tree specification. At the moment this idea of enlargement, used in Theorem 4, is being worked out to generalize the concept of rank correlation tree specifications to rank correlation graph specifications. In these latter specifications the graphs specifying the bivariate dependence relations may contain cycles. It seems that this generalization can be obtained by adding the restriction that the bivariate distributions have (rank) linear regression.

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.

van Dorp J.R. **(1991)**, *Dependence Modeling for Uncertainty Analysis*, Delft University of Technology, The Netherlands.

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

Halmos, P.R. **(1950)**, *Measure Theory*, D. Van Nostrand, Toronto.

Holland P.W. & Rosenbaum P.R. **(1986)**, *Conditional association and unidimensionality in monotone latent variable models*, The annals of statistics vol. 14, pp 1523-1543.

Hutchinson T.P. & Lai C.D. **(1990)**, *Continuous Bivariate Distributions, Emphasizing Applications*, Rumsby Scientific Publishing, Adelaide 5000, Australia.

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.

The following reports have appeared in this series:

94-16	Maurice Dohmen	Constraint techniques in interactive feature modeling
94-17	L.A. de Looff en E.W. Berghout	Vormen van onderzoek naar informatievoorziening in organisaties Geëffende paden bestaan niet
94-18	Henk J. Prins and Aad J. Hermans	Time domain calculations of the second-order drift force on a floating three-dimensional object in current and waves
94-19	W.A.J. Luxemburg, B. de Pagter, A.R. Schep	Diagonals of the powers of an operator on a Banach lattice
94-20	Ph. Clément, R. Hagmeijer, G. Sweers	On the invertibility of mappings arising in 2D grid generation problems
94-21	Florin Dan Barb en Willem de Koning	Het vaste-orde eindig-dimensionale dynamische digitale compensatieprobleem voor Pritchard-Salamon systemen
94-22	Serguei Foss, Gerard Hooghiemstra, Michael Keane	On a problem of Jon Wellner
94-23	T. Illés and G. Kassay	Farkas type theorems for generalized convexities
94-24	M. Zijlema, A. Segal and P. Wesseling	Finite volume computation of 2D incompressible turbulent flows in general coordinates on staggered grids
94-25	J.W. Boerstoel	A concept for the construction of grid-distributions mappings with a variational (adaptive) method for gradients of Jacobians of mappings
94-26	A. Lievaart	Studieprogramma voor tweede jaars wiskunde: een overzicht
94-27	E.W. Berghout en T.J.W. Renkema	Beoordelen van voorstellen voor investeringen in informatiesystemen: begrippenkader en vergelijking van methoden
94-28	Adrianus M.H. Meeuwissen and Roger M. Cooke	Tree dependent random variables

Copies of these reports may be obtained from the bureau of the Faculty of Technical Mathematics and Informatics, Julianalaan 132, 2628 BL DELFT, phone 015-784568.
A selection of these reports is available in PostScript form at the Faculty's anonymous ftp-site. They are located in the directory /pub/publications/tech-reports at ftp.twi.tudelft.nl.