Distribution - Free Continuous Bayesian Belief Nets

D. Kurowicka
Delft Institute for Applied Mathematics
Delft University of Technology
Delft
The Netherlands
D.Kurowicka@ewi.tudelft.nl

R.M.Cooke
Delft Institute for Applied Mathematics
Delft Univeristy of Technology
Delft
The Netherlands
R.M.Cooke@ewi.tudelft.nl

Abstract

This paper introduces distribution free continuous belief nets using the vine - copulae modelling approach. Nodes are associated with arbitrary continuous invertible distributions, influences are associated with (conditional) rank correlations and are realized by (conditional) copulae. Any copula which represents (conditional) independence as zero (conditional) correlation can be used. We illustrate this approach with a flight crew alertness model.

1 Introduction

Bayesian belief nets (bbns) are directed acyclic graphs representing high dimensional uncertainty distributions, and are becoming increasingly popular in modelling complex systems. Application is limited by the excessive assessment burden, which leads to informal or unstructured quantification. Continuous bbns exist only for the joint normal case, and require assessing means, conditional variances and partial regression coefficients (F.V. 2001)

This paper develops a "copula-free" approach to continuous bbns [1]. Any copula may be used so long as the chosen copula represents (conditional) independence as zero (conditional) correlation. This approach cannot rely on the equality of partial and conditional correlation, and hence cannot rely on vine transformations to deal with observation and updating, as done in (Kurowicka D. and Cooke R.M. 2002).

Nonetheless, it is shown that the elicitation protocol of (Kurowicka D. and Cooke R.M. 2002) based on conditional rank correlation can work in a copula-free environment. A unique joint distribution can be determined and sampled based on the protocol, which factorizes in the manner prescribed by the bbn and can be updated with observations of values or intervals. The theory presented here can be extended to include "ordinal" variables; that is variables which can be written as monotone transforms of uniform variables, perhaps taking finitely many values. The dependence structure must be defined with respect to the uniform variates. Further, we consider here only the case where the conditional correlations associated with the nodes of vines are constant; however, the sampling algorithms discussed below will work mutatis mutandis for conditional correlations depending on the values of the conditioning variables. We note that quantifying bbn's in this way requires assessing *all* (continuous, invertible) one dimensional marginal distributions. On the other hand, the dependence structure is meaningful for *any* such quantification. In fact, when comparing different decisions or assessing the value of different observations, it is frequently sufficient to observe the effects on the quantile functions of each node. For such comparisons we do not need to assess the one dimensional margins at all. This will be illustrated in the sequel.

We assume the reader is familiar with bbn's. We introduce vines and copula's in section 1, and distribution free bbn's in section 3. Section 4 illustrates their use with an example from (et. al. 2003) involving airline

---

[1]In (Kurowicka D. and Cooke R.M. 2002) the authors introduced an approach to continuous bbns using vines (Bedford T.J. and Cooke R.M. 2002) and the elliptical copula (Kurowicka D, Misiewicz J. and Cooke R.M. 2000). Influences were associated with conditional rank correlations, and these were realized by (conditional) elliptical copulae. While this approach has some attractive features, notably in preserving some relations between conditional and partial correlation, it also has disadvantages. Foremost among these is the fact that zero (conditional) correlation does not correspond to (conditional) independence under the elliptical copula.

flight crew alertness.

2 Vines and copulae

We build complex high dimensional distributions from two dimensional and conditional two dimensional distributions with uniform margins. The two dimensional distributions are called "copulae".

Definition 2.1 *A copula C is a distribution on the unit square with uniform margins. Random variables X and Y are joined by copula C if their joint distribution can be written*

$$F_{XY}(x, y) = C(F_X(x), F_Y(x)).$$

The diagonal band copula was introduced in (Cooke R.M., Waij R. 1986). For positive correlation the mass is concentrated on the diagonal band with vertical bandwidth $\beta = 1 - \alpha$. Mass is distributed uniformly on the rectangle with corners $(\beta, 0), (0, \beta), (1 - \beta, 1)$ and $(1, 1 - \beta)$ and is uniform but "twice as thick" in the triangular corners. The correlation coefficient is given by $\rho = sign(\alpha) \left((1 - |\alpha|)^3 - 2(1 - |\alpha|)^2 + 1 \right)$ (Cooke R.M., Waij R. 1986). For negative correlation the band is drawn between the other corners. Figure 1 shows the density of the diagonal band distribution with correlation 0.8.
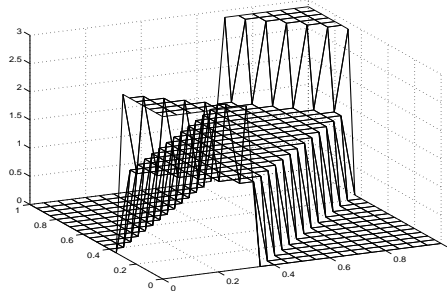

Figure 1: A density of the diagonal band copula with correlation 0.8.

In this paper we use the diagonal band copulae (or mixtures of these) because of their compliant analytical form. However we could use any copulae for which zero correlation entails independence, including the maximum entropy copulae, Frank's copulae, etc. For more information about copulae we refer to e.g. (H. 1997), (Doruet Mari D., Kotz S. 2001), (R.B. 1999) (D. 2004).

Graphical models called *vines* were introduced in (R.M. 1997; **?**). A vine on N variables is a nested set of trees $\{T_1, ...T_{n-1}\}$ where the edges of tree j are the nodes of tree j+1, and each tree has the maximum number of edges. A *regular vine* on N variables is a vine in which two edges in tree j are joined by an edge in tree j+1 only if these edges share a common node. A regular vine is called a *canonical vine* if each tree T_i has a unique node of degree N-i, hence has maximum degree. A regular vine is called a *D-vine* if all nodes in T_1 have degree not higher than 2 (see Figure 2). There are N(N-1)/2 edges in a regular vine on N variables. Each edge in a regular vine may be associated with a conditional a copula, that is, a conditional bivariate distribution with uniform margins (for j=1 the conditions are vacuous). The conditional bivariate distributions associated with each edge are determined as follows: the variables reachable from a given edge are called the *constraint set* of that edge. When two edges are joined by an edge of the next tree, the intersection of the respective constraint sets are the *conditioning variables*, and the symmetric differences of the constraint sets are the *conditioned variables*. The regularity condition insures that the symmetric difference of the constraint sets is always a doubleton. Each pair of variables occurs once as conditioned variables.

It is convenient to specify the conditional bivariate copulae by a *rank correlation specification*: first assign a constant conditional rank correlation to each edge of the vine, then choose a class of copulae indexed by

correlation coefficients in the interval $[-1, 1]$ and select the copulae with correlation corresponding to the conditional rank correlation. The density can be factorized into the product of bivariate densities depending on the conditioning variables of each edge (Bedford T.J. and Cooke R.M 2001) (see formula 2 below). For the precise definitions and properties of regular vines we refer to (Bedford T.J. and Cooke R.M. 2002; Kurowicka D. and Cooke R.M. 2003) A joint distribution satisfying the vine-copula specification can be constructed and sampled on the fly, and will preserve maximum entropy properties of the conditional bivariate distributions (R.M. 1997; Bedford T.J. and Cooke R.M 2001).
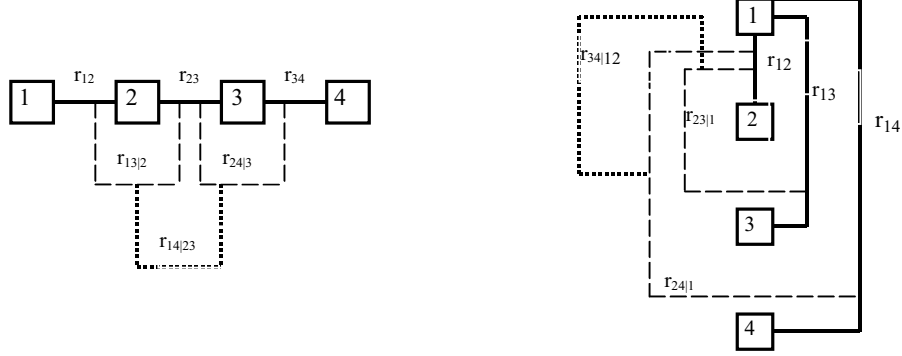

Figure 2: D-vine (left) and canonical vine (right) on 4 variables with (conditional) rank correlations assigned to the edges.

The rank correlation specification on a regular vine determines the whole joint distribution. There are two strategies for sampling such a distribution, which we term the *cumulative* and *density* approaches. We first illustrate the cumulative approach with the distribution specified by the D-vine in Figure 2, D(1,2,3,4): Sample four independent variables distributed uniformly on interval [0,1], U_1, U_2, U_3, U_4 and calculate values of correlated variables X_1, X_2, X_3, X_4 as follows:

1. $x_1 = u_1,$

2. $x_2 = F_{r_{12};x_1}^{-1}(u_2),$

3. $x_3 = F_{r_{23};x_2}^{-1}\left(F_{r_{13|2};F_{r_{12};x_2}(x_1)}^{-1}(u_3)\right),$

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

where where $F_{r_{ij|k};X_i}(X_j)$ denotes the cumulative distribution function for X_j, applied to X_j, given X_i under the conditional copula with correlation $r_{ij|k}$.

Notice that the sampling procedure for D-vine uses conditional distributions as well as inverse conditional distributions. We shorten the notation by dropping the "r"'s and write the *general sampling algorithm* as:

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

$$x_5 = F^{-1}_{5|4:x_4}\left(F^{-1}_{5|34:F_{3|4}(x_3)}\left(F^{-1}_{5|234:F_{2|34}(x_2)}\left(F^{-1}_{5|1234:F_{1|234}(x_1)}(u_5)\right)\right)\right),$$

$$\cdots$$

$$x_n = F^{-1}_{n|n-1:x_{n-1}}\left(F^{-1}_{n|n-2,n-1:F_{n-2|n-1}(x_{n-2})}\left(F^{-1}_{n|n-3,n-2,n-1:F_{n-3|n-2,n-1}(x_{n-3})}(\cdots\right.\right.$$
$$\left.\left.\left(F^{-1}_{n|1\ldots n-1:F_{1|2\ldots n-1}(x_1)}(u_n)\right)\cdots\right)\right);$$

When the bivariate distributions are indexed by conditional rank correlations, the correlations to be specified are

$$
\begin{array}{cccccc}
r_{12}, & r_{13|2}, & r_{14|23}, & r_{15|234}, & \cdots & r_{1,n-1|2\ldots n-2}, & r_{1,n|2\ldots n-1}, \\
& r_{23}, & r_{24|3}, & r_{25|34}, & \cdots & r_{2,n-1|3\ldots n-2}, & r_{2,n|3\ldots n-1}, \\
& & r_{34}, & r_{35|4}, & \cdots & r_{3,n-1|4\ldots n-2}, & r_{3,n|4\ldots n-1}, \\
& & & & \cdots \\
& & & & r_{n-2,n-1}, & r_{n-2,n|n-1}, \\
& & & & & r_{n-1,n},
\end{array}
\tag{1}
$$

Notice that the conditional rank correlations can be chosen arbitrarily in the interval $[-1, 1]$; they need not be positive definite or satisfy any further algebraic constraint. Two distributions with the same conditional correlations (1) and the same conditional copulae, are the same distribution.

When the vine-copula distribution is given as a density, the density approach to sampling may be used. Let $\mathcal{V} = (T_1 \ldots T_{n-1})$ be a regular vine on n uniform variables $(X_1, \ldots X_n)$, let E_m be the edge set for tree T_m, and for $e \in E_m$ with conditioning variables D_e, let $c_{ij|D_e}$ be the copula density associated with e. (Bedford T.J. and Cooke R.M 2001) show that the density for a distribution specified by the assignment of copulae to the edges of $\mathcal{V}$ is given by

$$\prod_{m=1}^{n-1}\prod_{e\in E_m} c_{ij|D_e}(F_{i|D_e}(x_i), F_{j|D_e}(x_j)) f_1(x_1)\ldots f_n(x_n) = \prod_{m=1}^{n-1}\prod_{e\in E_m} c_{ij|D_e}(F_{i|D_e}(x_i), F_{j|D_e}(x_j)) \tag{2}$$

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

This expression may be used to sample the vine distribution; namely, draw a large number of samples $(x_1, \ldots x_n)$ uniformly, and then resample these with probability proportional to (2). This is less efficient than the general sampling algorithm given previously; however it may be more convenient for conditionalization.

3 Continuous BBNs

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

1. Construct a sampling order for the nodes, that is, an ordering such that all ancestors of node i appear before i in the ordering. A sampling order begins with a source node and ends with a sink node. Of course the sampling order is not in general unique. Index the nodes according to the sampling order $1, \ldots, n$.

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

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

3. Underscore those nodes in each condition, which are not parents of the conditioned variable and thus are not necessary in sampling the conditioned variable. This uses (some of) the conditional independence

relations in the belief net. Hence if in sampling $2, \ldots, n$ variable 1 is not necessary (i.e. there is no influence from 1 to any other variable) then

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

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

4. For each term i with parents (non-underscored variables) $i_1..i_{p(i)}$ in (3), associate the arc $i_{p(i)-k} \longrightarrow i$ with the conditional rank correlation

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

$$r(i, i_{p(i)-k}|i_{p(i)}, \ldots i_{p(i)-k+1}); \ 1 \le k \le p(i) - 1. \tag{4}$$

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

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

Theorem 3.1 *Given a bbn with n nodes; the specification of conditional rank correlations (4), $i = 1, ...n$ and a copula realizing all correlations $[-1, 1]$ for which correlation 0 entails independence uniquely determines the joint distribution. This joint distribution satisfies the characteristic factorization (3) and the conditional rank correlations in (4) are algebraically independent.*

Proof. The first term in (3) is determined vacuously. We assume the joint distribution for $\{1, ...i - 1\}$ has been determined. Term i of the factorization (3) involves $i - 1$ conditional variables, of which $\{i_{p(i)+1}, ..i_{i-1}\}$ are conditionally independent of i given $\{i_1, ..i_{p(i)}\}$. We assign

$$r(i, i_j|i_1, \ldots i_{p(i)}) = 0; \ i_{p(i)} < i_j \le i - 1. \tag{5}$$

Then the conditional rank correlations (4, 5) are exactly those on $\mathcal{D}^i$ involving variable i, that is, in the last column of matrix (1). The other conditional bivariate distributions on $\mathcal{D}^i$ are already determined. It follows that the distribution on $\{1, ..i\}$ is uniquely determined. Since zero conditional rank correlation implies conditional independence,

$$P(1, ...i) = P(i|1, ...i - 1)P(1, ...i - 1) = P(i|i_1, ...i_{p(i)})P(i_1, ...i_{p(i)})$$

from which it follows that the factorization (3) holds.

Nodes and arcs can be added or deleted from a bbn quantified with this protocol, *without* re-assessing previously assessed correlations. This is significant difference with respect to parametric continuous bbns, in which partial regression coefficients of the child given all of its parents must be assessed. When a parent is added or deleted, the remaining coefficients must be assessed again.

We sample X_i using the sampling procedure for $\mathcal{D}^i$ In general it is not possible to keep the same order of variables in successive D-vine, and some conditional distributions will have to be calculated as in Example 3.1.

Example 3.1 *Let us consider the following bbn on 5 variables.*
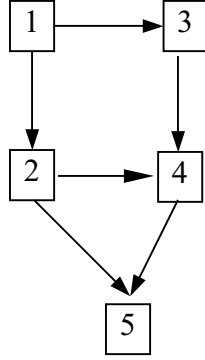
Sampling order: 1, 2, 3, 4, 5.

Factorization: $P(1)P(2|1)P(3|\underline{2}1)P(4|3\underline{2}\underline{1})P(5|4\underline{3}\underline{2}\underline{1})$.

Rank correlations that have to be assessed:

$$r_{21}, r_{31}, r_{43}, r_{42|3}, r_{54}, r_{52|4}.$$

In this case $\mathcal{D}^4 = D(4,3,2,1)$ but the order of variables in $\mathcal{D}^5$ must be $D(5,4,2,3,1)$. Hence this bbn cannot be represented as one vine.

Using the conditional independence properties of the bbn, the sampling procedure can be simplified as:

$x_1 = u_1,$

$x_2 = F^{-1}_{r_{21};x_1}(u_2),$

$x_3 = F^{-1}_{r_{31};x_1}(u_3),$

$x_4 = F^{-1}_{r_{43};x_3}\left(F^{-1}_{r_{42|3};F_{2|3}(x_2)}(u_4)\right),$

$x_5 = F^{-1}_{r_{54};x_4}\left(F^{-1}_{r_{52|4};F_{2|4}(x_2)}(u_5)\right).$

The conditional distributions $F_{2|3}(x_2), F_{2|4}(x_2)$ are not known and must be calculated:

$$
\begin{aligned}
f_{23}(x_2, x_3) &= \int_0^1 c_{21}(x_2, v)c_{13}(v, x_3)dv \\
F_{2|3}(x_2) &= \int_0^{x_2} f_{2|3}(u)du = \int_0^{x_2} f_{23}(u, x_3)du \\
F_{2|4}(x_2) &= \int_0^{x_2} \int_0^1 c_{43}(x_4, x_3) * f_{23}(x_2, x_3) * c_{42|3}\left(F_{4|3}(x_4), F_{2|3}(x_2)\right) dx_3
\end{aligned}
$$

where $c_{42|3}$ is a density of the copula with correlation $r_{42|3}$.

Since the diagonal band copula is not supported on the entire unit square, it is preferable to use mixtures of diagonal band copulae which realize the desired correlations while assigning positive probability to each point in the unit square. Good results are obtained by mixing a diagonal band copula with the independent copula; these mixtures have lower information relative to the uniform density, under a correlation constraint, and do not significantly impede computations. This was done in the following example.

4 Example: Flight crew alertness model

In Figure 3 a flight crew alertness model adapted from the discrete model described in (et. al. 2003) is presented. In the original model all chance nodes were discretized to take one of two values "OK" or "Not OK". Alertness is measured by performance by a simple tracking test during off-duty moments. The results are scored on an increasing scale and can be modelled as a continuous variable. Continuous distributions for each node must be gathered from existing data or expert judgement (R.M. 1991). The distribution functions are used to transform each variable to uniform on the interval (0,1). Required (conditional) rank correlations are found using the protocol described in Section 3. These can be assessed by experts in the way described in (B.C.P. 2002). In Figure 3 (conditional) rank correlation is assigned to each arc of the bbn. These numbers are chosen to illustrate this approach and are based on the assessments of in house experts.
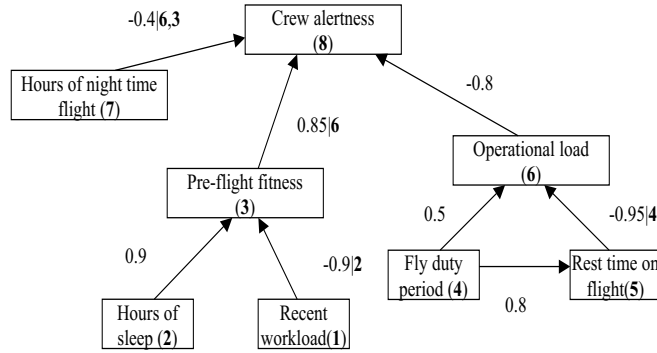

Figure 3: Flight crew alertness model.

The sampling algorithm for distribution described by the bbn in Figure 3 is the following.

$x_1 = u_1,\ x_2 = u_2,\ x_4 = u_4,\ x_7 = u_7,$

$x_3 = F^{-1}_{r_{32};x_2}(F^{-1}_{r_{13|2};x_1}(u_3)),$

$x_5 = F^{-1}_{r_{54};x_4}(u_5),$

$x_6 = F^{-1}_{r_{64};x_4}(F^{-1}_{r_{65|4};F_{5|4}(x_5)}(u_6)),$

$x_8 = F^{-1}_{r_{86};x_6}\left(F^{-1}_{r_{83|6|4};x_3}\left(F^{-1}_{r_{87|36};x_7}(u_8)\right)\right).$

The main use of bbns is in decision support, and in particular updating on the basis of possible observations. Let us suppose that we know before the flight that the crew didn't have enough sleep and they will have along flight. Let us assume that the crew's hours of sleep correspond to 25th percentile of hours of sleep distribution and the fly duty period is equal to 80th percentile of the flight duty period distribution.

We seek policies that could compensate loss of the crew alertness in this situation. Firstly we require that the number of night hours on the flight should be small (equal to 10th percentile). This improves situation a bit (dotted line in Figure 4). Alternatively we could require having long resting time on a flight (equal to 90th percentile). This results in a significant improvement of the crew alertness distribution (see dashed line in Figure 4). Combining these both polices improves the result even more.

Notice that in comparing different polices it is not necessary to know actual distributions of given variables. Our decisions can be based on quantile information. We might think of the transformation from
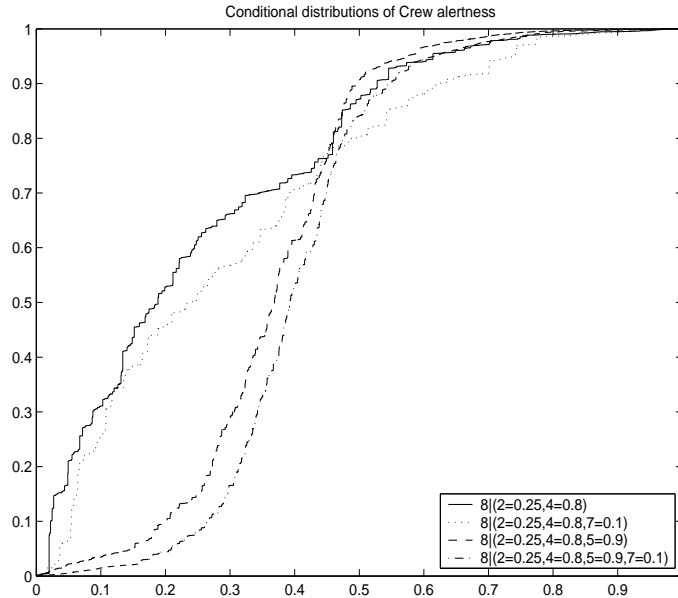
Figure 4: Four conditional distributions of crew alertness.

quantiles to physical units of the variables as being absorbed into a monotonic utility function. Thus, conclusions based on quantiles will hold for all monotonic utility functions of the random variables.

Notice also that this quantification requires eight numbers. If the individual nodes are described with discrete distributions involving K outcomes, then $22K$ algebraically independent numbers are required. This demonstrates the dramatic reduction of assessment burden obtained by quantifying influence as conditional rank correlation.

5   Conclusions

The discrete bbns have recently become a very popular tool in modelling of risk and reliability. Their popularity is based on the fact that influence diagrams capture engineer's intuitive understanding of complex systems, and at the same time serve as user interfaces for sophisticated software systems. Continuous bbns can significantly reduce the assessment burden. Parametric continuous bbns have the advantage of enabling analytic updating. On the other hand, assessing partial regression coefficients may be unintuitive, especially if the variables must first undergo transformation to joint normal. Further, adding or deleting variables requires re-assessing previously assessed partial regression coefficients. Distribution free continuous bbns have an advantage in this regard; their primary disadvantage is that updating must be done by Monte Carlo simulation.

References

B.C.P., K. (2002). *Probabilistic Inversion in Uncertainty Analysis and related topics.* Isbn 90-9015710-7, TU Delft.

Bedford T.J. and Cooke R.M (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Inteigence 32*, 245–268.

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

Cooke R.M., Waij R. (1986). Monte carlo sampling for generalized knowledge dependence with application to human reliability. *Risk Analysis 6*, 335–343.

D., L. (2004). Generalized diagonal band copulae - applications to the vine–copula method. Presented at the DeMoSTAFI 2004 conference in Quebec, Canada.

Doruet Mari D., Kotz S. (2001). *Correlation and Dependence*. London: Imperial College Press.

et. al., R. A. (2003). Casual modeling for integrated safety at airport. *Proc ESREL 2003, Safety and Reliability 2*, 1321–1327.

F.V., J. (2001). *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag.

H., J. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Kurowicka D. and Cooke R.M. (2002). The vine copula method for representing high dimensional dependent distributions; application to continuous belief nets. *Proc. Winter Simulation Conference*.

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

Kurowicka D, Misiewicz J. and Cooke R.M. (2000). Elliptical copulae. *Proc of the International Conference on Monte Carlo Simulation - Monte Carlo*, 209–214.

R.B., N. (1999). *An Introduction to Copulas*. New York: Springer.

R.M., C. (1991). *Experts in Uncertainty*. New York: Oxford University Press.

R.M., C. (1997). Markov and entropy properties of tree and vines- dependent variables. In *Proceedings of the ASA Section of Bayesian Statistical Science*.