

Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets

A.M.Hanea, D.Kurowicka, R.M.Cooke

Delft Institute of Applied Mathematics
Delft University of Technology
The Netherlands

P.van Leeuwen

National Aerospace Laboratory
Amsterdam
The Netherlands

Abstract

Bayesian Belief Nets (BBNs) have become a very popular tool to specify high dimensional probabilistic models. Their popularity is based on the fact that influence diagrams capture an engineer's intuitive understanding of complex systems. Commercial tools with an advanced graphical user interface that support BBNs construction and inference are available. Thus, building and working with BBNs is very efficient as long as one is not forced to quantify complex BBNs. A high assessment burden of discrete BBNs is often caused by the discretization of continuous random variables. Until recently, continuous BBNs were restricted to the joint normal distribution. In [1] the 'copula - vine' approach to continuous BBNs is presented. This approach is quite general and allows traceable and defensible quantification methods, but it comes at a price: these BBNs must be evaluated by Monte Carlo simulation. Updating such a BBN will require re-sampling the whole structure. The advantages of fast updating algorithms for discrete BBNs are decisive. In this paper we combine the reduced assessment burden and modelling flexibility of the continuous BBNs with the fast updating algorithms of discrete BBNs.

- Quantify nodes of a BBN as continuous univariate random variables and arcs as conditional rank correlations
- Sample this structure
- Use the sample file in Netica to build conditional probability tables for a discretized version of the BBN
- Perform fast updating.

We will address some computational problems of this approach, as well as propose ways to solve them. We illustrate it with some practical examples.

Keywords: bayesian belief nets, dependence modelling, vines, multivariate probability distribution

1 Introduction

A Bayesian Belief Net (BBN) is a directed acyclic graph, together with an associated set of probability tables. The nodes of the graph represent

random variables, which can be discrete or continuous, and the arcs represent causal relationships between variables. BBNs enable us to model high dimensional uncertainty distributions. The visual representation can be very useful in clarifying previously opaque assumptions about the dependencies between different variables. Applications involving high complexity in data-sparse environments are severely limited by the excessive assessment burden which leads to rapid, informal and indefensible quantification. This assessment burden can only be reduced by a drastic discretization of the chance nodes. In many cases continuous nodes would be more appropriate.

Until recently, continuous BBNs were restricted to the joint normal distribution. In [1] the authors introduced an approach to continuous BBNs using vines [2] together with copulas that have the *zero independence property*¹. In the procedure proposed here, nodes are associated with arbitrary continuous invertible distributions and arcs with conditional rank correlations, which are realized by the chosen copula. In order to quantify BBNs using the copula-vine approach, one needs to specify all one dimensional marginal distributions and the (conditional) rank correlations.

The rank correlation specification on a vine plus the choice of a copula with the zero independence property, determines the whole joint distribution. As presented in [1], the joint distribution is factorized in the standard way (for BBNs) following a (non - unique) sampling order for the nodes. For each term of the factorization, a D-vine is built. The sampling algorithm consists of sampling the i^{th} variable in the ordering according to the i^{th} D-vine. The sampling procedure for

¹Section 2 of the paper presents a number of definitions and preliminary results, explaining this property among other things.

D-vines is fully described in [3]. In the D-vines used, the order of the variables might change, hence in general it is not possible to represent a BBN with just one D-vine (see section 3).

In most cases, where more than one vine is needed, some conditional distributions have to be calculated. These calculations consist of evaluating multiple integrals by Monte Carlo simulations. For accurate results, a very large number of samples is needed in the sampling procedure. For each of those samples, one will need to calculate the numerical value of the multiple integrals. In case of a large number of variables, one would have to be prepared to run the model for a few days. Moreover, updating such a BBN will require re-sampling the whole structure every time new evidence becomes available. In these cases the advantages of fast updating algorithms for discrete BBNs are decisive. This motivates the approach advanced in section 4 of this paper, which consists of combining the reduced assessment burden and modelling flexibility of the continuous BBNs with the fast updating algorithms of discrete BBNs. This can be done, using vine sampling together with existing discrete BBNs software. Even though this procedure has some attractive features, remarkably in the fact that the structure needs to be sampled just once, it also has disadvantages. We will present a small example in which sampling just this one time becomes very complicated and time consuming under certain conditions.

To overcome this problem, a new method for sampling the BBN is proposed in section 5. The influences will still be represented as (conditional) rank correlations specified on D-vines. We will transform the variables to standard normals and the (conditional) rank correlations to (conditional) product moment correlations via Pearson's formula ([5]). Using the properties of the multivariate normal distribution, we now have a partial correlation vine specification that uniquely determines the correlation matrix, as shown in [2]. Further, we can induce this product moment correlation structure using well known methods, and transform back to the original variables.

In doing so, the possible change in order of the variables in the D-vines does not produce any significant inconvenience. Going from one vine to another requires only calculating new partial correlations. However, these calculations are not

very time consuming, so the computational time is reduced considerably. Examples and comparisons will be presented at the end of this paper.

2 Definitions & Concepts

We will assume that the reader is acquainted with the theory about discrete and gaussian BBNs.

Definition 2.1. *The copula of two continuous random variables X and Y is the joint distribution of $(F_X(X), F_Y(Y))$, where F_X, F_Y are the cumulative distribution functions of X, Y respectively.*

The copula of (X, Y) is a distribution on $[0, 1]^2$ with uniform marginal distributions.

Definition 2.2. *A copula is said to have the zero independence property if zero correlation entails the independent copula.*

A graphical model called vines was introduced in [2]. A vine on n variables is a nested set of trees. The edges of the j^{th} tree are the nodes of the $(j + 1)^{th}$ tree. Each tree in a vine has maximum number of edges.

Definition 2.3. \mathcal{V} is called a regular vine on n elements if:

1. $\mathcal{V} = (T_1, \dots, T_n)$;
2. T_1 is a connected tree with nodes $N_1 = 1, \dots, n$, and edges E_1 and for $i = 2, \dots, n-1$ T_i is a tree with nodes $N_i = E_{i-1}$;
3. For $i = 2, \dots, n-1$, $a, b \in E_i$, $\#a \Delta b = 2$, where Δ denotes the symmetric difference.

Definition 2.4. A regular vine is called a:

- **D-vine** if each node in T_1 has the degree at most 2 (see Figure 2.1);
- **C-vine** if each tree T_i has a unique node of degree $n - i$. The node with maximal degree in T_1 is called the root (see Figure 2.2).

Each edge of a regular vine may be associated with a constant conditional rank correlation which can be arbitrarily chosen in the interval $[-1, 1]$. Using a copula to realize these rank correlations, a joint distribution satisfying the copula-vine specification can be constructed and it will always be consistent. For rigorous definitions and

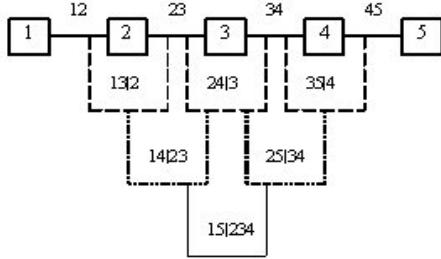


Figure 2.1: D-vine on 5 variables.

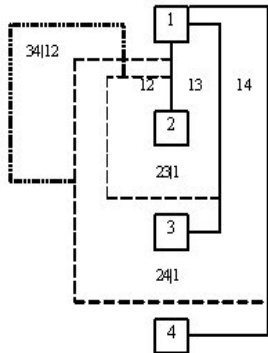


Figure 2.2: C-vine on 4 variables.

proofs we refer to [3]. Each vine branch may also be associated with partial correlations. First, we will recall the definition of the partial correlation.

Definition 2.5. [4] Let $X_1 \dots X_n$ be random variables. The partial correlation of X_1 and X_2 given $X_3 \dots X_n$ is:

$$\rho_{12|3,\dots,n} = \frac{\rho_{12|4,\dots,n} - \rho_{13|4,\dots,n} \cdot \rho_{23|4,\dots,n}}{((1 - \rho_{13|4,\dots,n}^2) \cdot (1 - \rho_{23|4,\dots,n}^2))^{\frac{1}{2}}}$$

Partial correlations can be computed from correlations using the above recursive formula.

Definition 2.6. A complete partial correlation vine specification is a regular vine with a partial correlation specified for each edge.

One can remark that the definition above does not have any reference to a copula as, in general, the partial correlation is not a property of a copula. However, this is not the case for the bivariate normal distribution. For joint normal variables the partial correlation is equal to the conditional product moment correlation, so one

can define a complete normal partial correlation vine specification. In [2] it is shown how the notion of a regular vine can be used to construct a joint normal distribution.

Theorem 2.1. Given any complete partial correlation vine specification for standard normal random variables X_1, \dots, X_n , there is a unique joint normal distribution for X_1, \dots, X_n satisfying all partial correlation specifications.

A very important result from [2] is that each partial correlation vine specification uniquely determines the correlation matrix, even without the assumption of joint normality.

Theorem 2.2. For any regular vine on n elements there is a one to one correspondence between the set of $n \times n$ positive definite correlation matrices and the set of partial correlation specification for the vine.

In the end of this section we will summarize some well-known facts about the multivariate normal distribution.

If (X, Y) has the bivariate normal distribution, a necessary and sufficient condition for X and Y to be independent is that $\rho(X, Y) = 0$.

In the joint normal distribution's case, one can find the relationship between product moment correlation (ρ) and rank correlation (r) using Pearson's transformation.

Proposition 2.1. [5] Let (X, Y) be a random vector with the joint normal distribution, then:

$$\rho(X, Y) = 2 \sin\left(\frac{\pi}{6} \cdot r(X, Y)\right).$$

Let X be a n -dimensional random vector with multivariate normal distribution. Let the vector μ be the expected value of X , and V be its covariance matrix. For a fixed $k < n$ consider the partition of X , μ and V given below:

$$X = \begin{pmatrix} X_a \\ X_b \end{pmatrix}, \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix}$$

where

$$X_a = (X_1, \dots, X_k)', X_b = (X_{k+1}, \dots, X_n)', \\ \mu_a = (\mu_1, \dots, \mu_k)', \mu_b = (\mu_{k+1}, \dots, \mu_n)', \\ V_{ii} = \text{var}(X_{i \in \{a,b\}}) \text{ and } V_{ab} = \text{cov}(X_a, X_b).$$

The conditional variance of X_b given X_a is denoted by $\text{var}_{b|a}(X_b)$.

Proposition 2.2. [6] (Marginal and conditional density function): If the partitioned random vector follows the distribution:

$$(X_a, X_b) \sim N \left[(\mu_a, \mu_b), V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix} \right],$$

then:

(i) the marginal distribution of X_a is normal with mean μ_a and variance V_{aa} ;

(ii) the conditional distribution of $(X_b|X_a)$ is normal with the mean:

$$E_{b|a}(X_b) = \mu_b + B_{b|a} \cdot (x_a - \mu_a),$$

where

$$B_{b|a} = V_{ba}V_{aa}^{-1};$$

and the variance:

$$\text{var}_{b|a}(X_b) = V_{bb|a} = V_{bb} - V_{ba}V_{aa}^{-1}V_{ab}.$$

3 Continuous BBNs & Vines

Having put all definitions and concepts in place, let us now consider a non-parametric continuous BBN. As we already said, in this case, the nodes are continuous univariate random variables and the arcs are associated with (conditional) parent-child rank correlations. We assume throughout that all univariate distributions have been transformed to uniforms on $(0, 1)$. The high dimensional joint distribution is specified using the vinecopula approach. Any copula may be used as long as it represents independence as zero correlation. The conditional rank correlations are algebraically independent, and there are tested protocols for their use in structured expert judgement (these protocols are presented in Chapter 2 of [3]). One can assign rank correlations to the arcs of a BBN according to the protocol presented in [1]. We will illustrate this procedure with an example.

Example 3.1. Let us consider the BBN in the Figure 3.1.

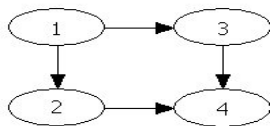


Figure 3.1: BBN with 4 nodes and 4 arcs.

There are two sampling orders for this structure: 1, 2, 3, 4, or 1, 3, 2, 4. Let us choose 1, 2, 3, 4.

The factorization is:

$$P(1)P(2|1)P(3|21)P(4|321).$$

We recall that the underscored nodes are the ones which are not necessary in sampling the conditioned variable. Hence, the (conditional) correlations that need to be assigned to the edges of this BBN are²: $\{r_{21}, r_{31}, r_{42}, r_{43|2}\}$. For each term i of the factorization ($i = 1, \dots, 4$), a D-vine on i variables is built. This D-vine is denoted by \mathcal{D}^i and it contains: the variable i , the non-underscored variables, and the underscored ones, in this order. Figure 3.2 shows the D-vines on variables 2, 3 and 4.

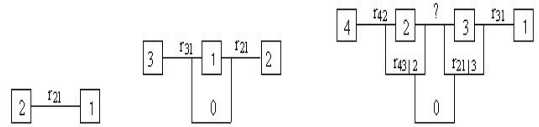


Figure 3.2: $\mathcal{D}^2, \mathcal{D}^3, \mathcal{D}^4$ for Example 3.1.

In the figure above, one can notice that $\mathcal{D}^3 = D(3, 1, 2)$, and the order of the variables from \mathcal{D}^4 must be $D(4, 3, 2, 1)$. Hence, this BBN cannot be represented as just one D-vine. In this case the sampling algorithm consists of sampling the i^{th} variable in the ordering according to \mathcal{D}^i , and calculating some conditional distributions. The procedure, which is presented below, starts with sampling four independent, uniform $(0, 1)$ variables, say U_1, \dots, U_4 .

$$\begin{aligned} x_1 &= u_1; \\ x_2 &= F_{r_{21}; x_1}^{-1}(u_2); \\ x_3 &= F_{r_{31}; x_1}^{-1}(F_{r_{32|1}; F_{r_{21}; x_1}(x_2)}^{-1}(u_3)); \\ x_4 &= F_{r_{42}; x_2}^{-1}(F_{r_{43|2}; F_{r_{32}; x_2}(x_3)}^{-1}(\\ &\quad F_{r_{41|32}; F_{r_{21|3}; F_{r_{32}; x_3}(x_2)}^{-1}(F_{r_{31}; x_3}(x_1))^{-1}(u_4))). \end{aligned}$$

The BBN structure reads the conditional independence of the variables X_3 and X_2 given X_1 , therefore $r_{32|1} = 0$, which gives:

$$F_{r_{32|1}; F_{r_{21}; x_1}(x_2)}^{-1}(u_3) = u_3,$$

Moreover, X_4 and X_1 are independent given X_2 and X_3 , hence $r_{41|32} = 0$, and so

²One could just as well specify $\{r_{21}, r_{31}, r_{43}, r_{42|3}\}$ instead.

$$F_{r_{41|32};F_{r_{21|3};F_{r_{32};x_3}(x_2)}(F_{r_{31};x_3}(x_1))(u_4) = u_4.$$

Consequently, using the conditional independence properties of the BBN, the sampling procedure can be simplified as:

$$\begin{aligned} x_1 &= u_1; \\ x_2 &= F_{r_{21};x_1}^{-1}(u_2); \\ x_3 &= F_{r_{31};x_1}^{-1}(u_3); \\ x_4 &= F_{r_{42};x_2}^{-1}(F_{r_{43|2};F_{r_{32};x_2}(x_3)}^{-1}(u_4)). \end{aligned}$$

The conditional distribution $F_{r_{32};x_2}(x_3)$ is not given³, so it must be found by calculating:

$$F_{3|2}(x_3) = \int_0^{x_3} \int_0^1 c_{21}(x_2, x_1) c_{31}(v, x_1) dx_1 dv,$$

where c_{i1} is the density of the chosen copula (we use Frank's copula [9]) with correlation r_{i1} , $i \in \{2, 3\}$.

For each sample, one will need to calculate the numerical value of the double integral⁴. In this case, when only one double integral needs to be evaluated, it can be easily done without excessive computational burden.

If some of the variables become certain, the results of sampling this model - conditional on their values - are obtained either by sampling again the structure (cumulative approach), or by using the density approach. We will present both methods in short, and for details we refer to [3].

Let us assume we learn $X_2 = 0.85$. In the cumulative approach the sampling procedure becomes:

$$\begin{aligned} x_1 &= F_{r_{21};x_2}^{-1}(u_1); \\ x_2 &= 0.85; \\ x_3 &= F_{r_{31};x_1}^{-1}(u_3); \\ x_4 &= F_{r_{42};x_2}^{-1}(F_{r_{43|2};F_{r_{32};x_2}(x_3)}^{-1}(u_4)). \end{aligned}$$

In the density approach, the conditionalization is made using the joint density, which can be evaluated as follows ([2]):

$$g(x_1, \dots, x_8) = c_{21}(x_2, x_1) c_{31}(x_3, x_1) c_{42}(x_4, x_2) c_{43|2}(F_{4|2}(x_4), F_{3|2}(x_3)).$$

Whichever of the two methods is preferred, the double integral still needs to be evaluated for

³We recall that this represents the conditional distribution $F_{3|2}(x_3)$.

⁴All numerical results in this paper are obtained using Matlab.

each sample, and for any new policy.

If the BBN consists of a cycle⁵ of 5 variables, and the same sampling procedure is applied, a triple integral will have to be calculated. The bigger the cycle is, the larger the number of multiple integrals that have to be numerically evaluated. And yet, this is not the worst that can happen⁶; an example of such a situation will be presented in section 5 of this paper.

The BBNs that resemble real life problems will often be quite big, and may well contain cycles of five or more variables. Updating such a structure is done by re-sampling the network each time a new policy is evaluated. In the case of a large number of variables, one would have to be prepared to run the model for a few days. To overcome this limitation we would like to combine the vine approach to the continuous BBNs, with the benefits of the discrete BBNs software. This is done in section 4 below.

4 Updating with Netica

Sampling a large BBN structure every time new evidence becomes available does not seem a very good idea in terms of computational time. On the other hand, sampling it just once, and employ the easiness of use, flexibility, good visualization, and fast updating of a commercial BBN tool, provides an elegant solution to this problem.

In doing so, we have chosen the free demo version (full-featured but limited in the size of the model) of Netica⁷.

The hybrid method proposed can be summarized as follows:

1. Quantify nodes of a BBN as continuous univariate random variables and arcs as parent-child rank correlations;
2. Sample this structure creating a large sample file;
3. Use this sample file (in Netica) to build conditional probability tables for a discretized version of the continuous BBN;

⁵Whenever we speak of cycles, we mean undirected cycles.

⁶More examples of BBN structures in which additional numerical calculations are needed are presented in Chapter 6 of [3].

⁷<http://www.norsys.com/>

- Use Netica to visualize the network and perform fast updating for the discretized BBN.

In order to perform the 3^{rd} step, a network has to be pre-prepared in Netica. This will contain the nodes of the BBN, each discretized in a certain - not necessarily small - number of states, together with the connections. Note that nothing will be filed into the (conditional) probability tables corresponding to each variable. We will illustrate this method by means of an extensive example.

Example 4.1. Flight Crew Alertness

In Figure 4.1, the so-called flight crew alertness model is given. A discrete form of this model was first presented in [7] and an adapted version of it was discussed in [1]. Figure 4.1 resembles the latest version of the model. The variables are continuous and their distribution functions are used to transform the variables to uniforms on (0, 1). The (conditional) rank correlations assigned to each arc of the BBN are chosen by the authors of [1] for illustrative purposes.

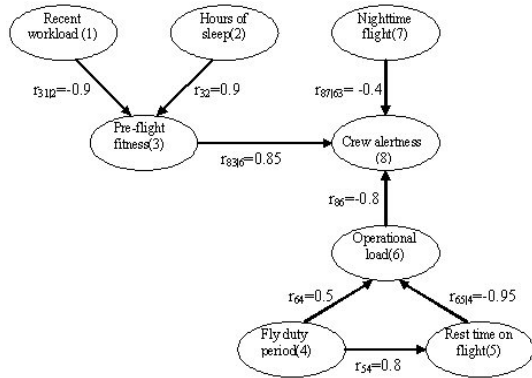


Figure 4.1: Flight crew alertness model. A number is assigned to each variable (on the right hand side of each name).

The sampling order is 1, 2, 3, 4, 5, 6, 7, 8. The sampling procedure for this structure uses Frank’s copula, and does not require any additional calculations:

$$\begin{aligned}
 x_1 &= u_1 \\
 x_2 &= u_2 \\
 x_3 &= F_{r_{32};x_2}^{-1}(F_{r_{31};x_1}^{-1}(u_3)) \\
 x_4 &= u_4 \\
 x_5 &= F_{r_{54};x_4}^{-1}(u_5)
 \end{aligned}$$

$$\begin{aligned}
 x_6 &= F_{r_{64};x_4}^{-1}(F_{r_{65};x_4}^{-1}(F_{r_{54};x_4}^{-1}(u_6))) \\
 x_7 &= u_7 \\
 x_8 &= F_{r_{86};x_6}^{-1}(F_{r_{83};x_6}^{-1}(F_{r_{87};x_7}^{-1}(u_8)))
 \end{aligned}$$

Figure 4.2 below shows the BBN from example 4.1, modelled in Netica. The variables are uniform on the (0, 1) interval, and each is discretized in ten states. A case file containing $8 \cdot 10^5$ samples, obtained using the sampling procedure described, was imported in Netica via the option ”Relation/Incorporate Case File”. This automatically creates the conditional probability tables needed for performing inferences.

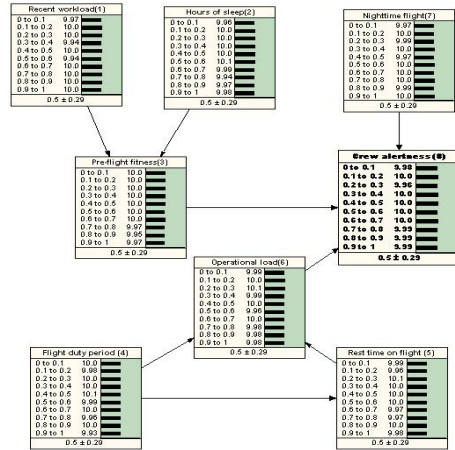


Figure 4.2: Flight crew alertness model in Netica.

Not only is the degree of realism greater in the continuous model, but also the quantification of the discretized BBN would require 12,140 probabilities, whereas the quantification with continuous nodes requires only 8 algebraically independent conditional rank correlations.

Figures 4.3, 4.4, and 4.5 present the conditional distribution of the variable ”Flight crew alertness(8)” given certain values of the variables ”Hour of sleep(2)”, and ”Fly duty period(4)”, obtained in two ways:

- using the vines-Netica updating;
- using the vines updating with the density approach.

After the sample file is imported in Netica, we conditionalize on ”Hours of sleep” $\in [0.2, 0.3]$ and ”Fly duty period” $\in [0.8, 0.9]$. We can use Netica to generate samples from the conditional distribution of ”Crew alertness”. In the same manner,

we sample from "Hours of sleep" $\in [0.2, 0.3]$ and "Fly duty period" $\in [0.8, 0.9]$ and save the samples that Netica generates via the option "Network/Simulate Cases". In the Monte-Carlo simulation for vines updating, we will have to re-sample the structure, in the same conditions ("Hours of sleep" $\in [0.2, 0.3]$; "Fly duty period" $\in [0.8, 0.9]$). For better results of the comparisons, we use the samples that we saved from Netica, in the simulation for updating with vines.

In Figure 4.3, the conditional probability tables from Netica were built using 10^4 samples. The agreement between the two methods is very poor, and this is because there are 10^3 different input vectors for node 8, each requiring 10 probabilities for the distribution of 8 given the input. With 10^4 samples, we expect each of the 10^3 different inputs to occur 10 times, and we expect a distribution on 10 outcomes to be very poorly estimated with 10 samples. Moreover, updating with vines does not produce a very smooth and accurate curve, also because the simulation was performed with 10^4 samples.

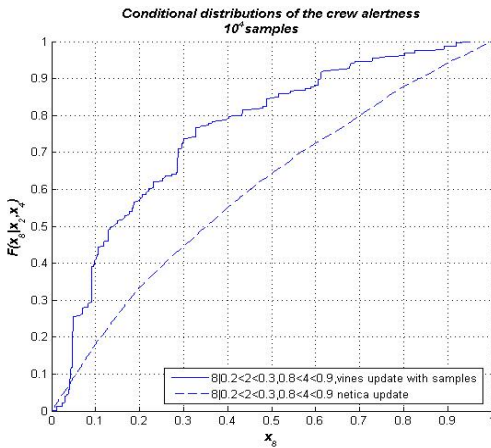


Figure 4.3: Distribution of $X_8|X_2, X_4$. Comparison of updating results in vines and Netica using 10^4 samples.

In the next figure, the sample file imported in Netica contains $8 \cdot 10^5$ samples which allows a very good estimation of the conditional distribution of "Crew alertness". Another 10^4 samples for "Hours of sleep" $\in [0.2, 0.3]$ and 10^4 for "Fly duty period" $\in [0.8, 0.9]$ are saved from Netica and used in the vines updating. The curves start to look

very similar indeed, but the one corresponding to vines (alone) updating is still not smooth because of the number of samples.

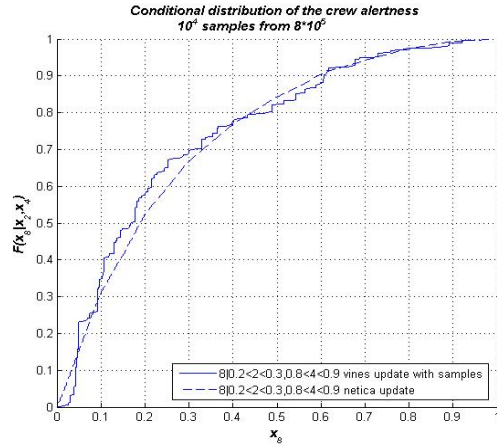


Figure 4.4: Distribution of $X_8|X_2, X_4$. Comparison of updating results in vines and Netica using 10^4 from $8 \cdot 10^5$ samples.

If we now do everything with the entire sample file of $8 \cdot 10^5$ samples, the agreement between the two conditional distributions is impeccable (see Figure 4.5). This motivates the use of a very big sample file.

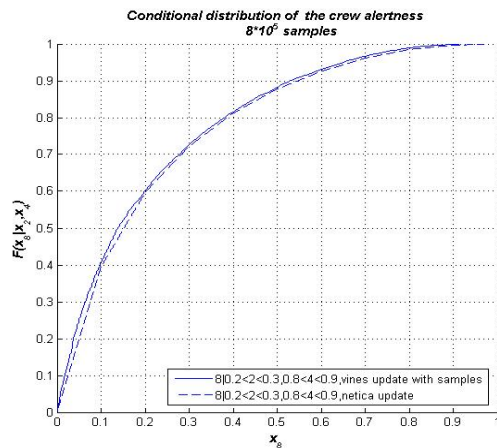


Figure 4.5: Distribution of $X_8|X_2, X_4$. Comparison of updating results in vines and Netica using $8 \cdot 10^5$ samples.

For a BBN with nodes that require a large number of inputs (large number of parent nodes,

discretized in fairly many states) the sample files should also be very large. The big advantage is that this huge sample file needs to be done once. Note however that in some cases it might happen that sampling the structure, even just once will cause problems, as we already mentioned in Section 3.

Example 4.2. Let us consider a BBN structure, which at a first glance, seems very easy to deal with, in the sense that it offers a lot of information about the dependence structure. This is presented in Figure 4.6.

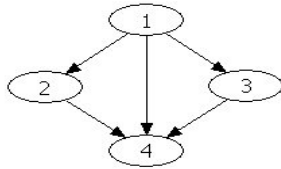


Figure 4.6: BBN with 4 nodes and 5 arcs.

If the set of (conditional) rank correlations that can be elicited is either $\{r_{21}, r_{31}, r_{42}, r_{41|2}, r_{43|21}\}$, or $\{r_{21}, r_{31}, r_{43}, r_{41|3}, r_{42|31}\}$, then the BBN can be represented as one D-vine, and so the sampling procedure does not require any extra calculations.

If, for some reason, these rank correlations cannot be specified, and the only ones which are available are: $\{r_{21}, r_{31}, r_{43}, r_{42|3}, r_{41|32}\}$ the situation worsens considerably.

The BBN can no longer be represented as one D-vine, since the order of the variables in \mathcal{D}^3 is 3, 1, 2, and in \mathcal{D}^4 is 4, 3, 2, 1. To sample X_4 , one needs to calculate:

$$x_4 = F_{4|3:x_3}^{-1} (F_{4|23:F_{2|3}(x_2)}^{-1} (F_{4|123:F_{1|23}(x_1)}^{-1} (u_4))).$$

The conditional distribution $F_{2|3}(x_2)$, can be found by evaluating a similar double integral as in Example 3.1. Furthermore, $F_{1|23}(x_1)$ needs to be calculated. This is, in fact, the conditional distribution of $F_{1|2}(x_1)$, given $F_{3|2}(x_3)$. Even though all the information needed seems to be available, evaluating the joint distribution of these two quantile functions turns out to be a very difficult job. Moreover at each step of its evaluation, one

should calculate the numerical value of the double integral for $F_{3|2}(x_3)$. This is a task that takes time and patience.

If this kind of calculations are necessary for such a small BBN, it is very likely that more complicated ones will be involved in larger structures. The time spent to solve this sort of problems would be, by far, much longer than one can afford.

5 Joint Normal Vines

All the troubles discussed until now are caused by the different sampling order of variables from one vine to another. To avoid these problems we advance here a new way of realizing the rank correlation specification on a regular vine using the joint normal distribution instead of a copula.

Let us start with a rank correlation vine specification on the variables X_1, \dots, X_n , with continuous, invertible distribution functions F_1, \dots, F_n . We adopt the following protocol:

1. Transform X_1, \dots, X_n to the standard normal variables Y_1, \dots, Y_n via the transformation $Y_i = \Phi^{-1}(F_i(X_i))$, ($\forall i$) ($i = 1, \dots, n$), where Φ is the cumulative distribution function of the standard normal distribution.
2. Construct the vine for the standard normal variables Y_1, \dots, Y_n . Since $\Phi^{-1}(F_i(X_i))$ are strictly increasing transformations, we assign the same (conditional) rank correlations to the edges of this vine.
3. To each edge of this vine assign $\rho_{i,j|D} = 2 \sin(\frac{\pi}{6} \cdot r_{i,j|D})$, where $\{i, j\}$ and D are the conditioned and conditioning sets, respectively, of the edge, and $r_{i,j|D}$ is the conditional correlation assigned to the corresponding edge from the initial vine. We now have a complete partial correlation vine specification⁸ for Y_1, \dots, Y_n . Theorem 2.1 ensures that there is a unique joint normal distribution for Y_1, \dots, Y_n satisfying all partial correlation specifications. Moreover there is an unique correlation matrix determined by this vine (Theorem 2.2).
4. Compute the correlation matrix R using the recursive formula from Definition 2.5.

⁸Conditional and partial correlations are equal for normal variables.

5. Sample the joint normal distribution of Y_1, \dots, Y_n , with correlation matrix R . ([8])
6. For each sample, calculate:

$$(F_1^{-1}(\Phi(y_1^j)), F_2^{-1}(\Phi(y_2^j)), \dots, F_n^{-1}(\Phi(y_n^j))),$$

where $((y_1^j), (y_2^j), \dots, (y_n^j))$ is the j^{th} sample from the previous step.

In this way we realize the joint distribution of the initial variables X_1, \dots, X_n , together with the dependence structure specified.

In case of a BBN which cannot be represented as one vine, we can make use of the protocol described above. We calculate everything on the joint normal vine, hence we can reorder the variables and recompute all partial correlations needed.

Further, we will present comparisons between the "joint normal vine" method and the copula-vine method together with Netica updating, using the BBN from Example 3.1. The marginal distributions of X_1, X_2, X_3, X_4 are uniform on the interval $(0, 1)$. We sample the structure both with the copula-vine approach, and the joint normal vine approach. Hence, we produce two sample files, each containing 10^5 samples. The resulting files are imported in Netica, and conditioning is performed in both cases.

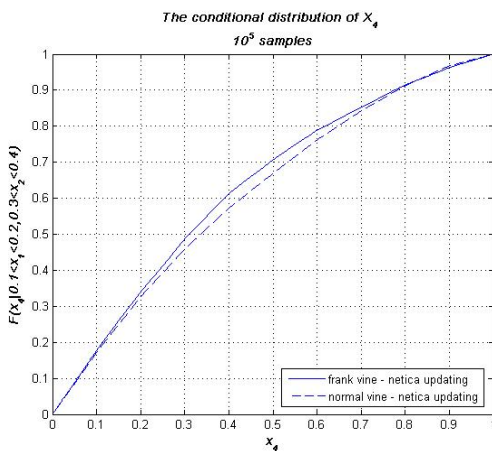


Figure 5.1: The distribution of $X_4|X_1, X_2$. Frank's Copula Vine vs Joint Normal Vine (conditioning in Netica using 10^5 samples).

Figure 5.1 presents the conditional distribution of the variable X_4 given that $X_1 \in [0.1, 0.2]$ and $X_2 \in [0.3, 0.4]$, obtained using the sample files produced with the two methods. One can notice a small disagreement between the two conditional distributions. If we think of this joint normal method in terms of the copula-vine method, where we made use of the normal copula, we can say that the difference between the two conditional distributions from Figure 5.1 is due to the different choice of copula.

Another way of comparing these methods is to calculate and compare the two sample correlation matrices. The matrix presented below corresponds to the sample file obtained using the copula-vine approach:

$$\begin{pmatrix} 1 & 0.4031 & 0.7028 & 0.3746 \\ 0.4031 & 1 & 0.2843 & 0.2028 \\ 0.7028 & 0.2843 & 1 & 0.5201 \\ 0.3746 & 0.2028 & 0.5201 & 1 \end{pmatrix}$$

The next matrix is obtained using the sample file generated with the joint normal vine method:

$$\begin{pmatrix} 1 & 0.4000 & 0.6974 & 0.3843 \\ 0.4000 & 1 & 0.2837 & 0.1985 \\ 0.6974 & 0.2837 & 1 & 0.5271 \\ 0.3843 & 0.1985 & 0.5271 & 1 \end{pmatrix}$$

Comparing the two matrices one can observe differences of order 10^{-3} , which represent a reasonable result taking into account the sampling errors.

The big advantage of this method is that the simulation runs for a few seconds, whereas with the previous sampling algorithm (in which additional calculations were necessary) the results were available in hours.

The same kind of results we find when we examine the structure from Example 4.1.

Figure 5.2 shows the conditional distribution of the variable "Flight crew alertness(8)" given the variables "Hour of sleep(2)" and "Fly duty period(4)". We can again notice that the choice of copula produces a small discrepancy between the curves. Comparing the two sample correlation matrices for this example we find that the maximum difference is $8 \cdot 10^{-3}$.

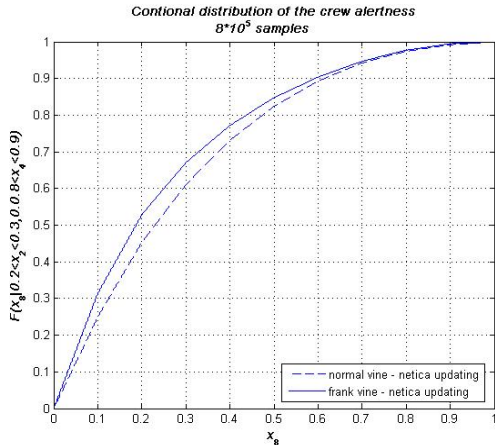


Figure 5.2: The distribution of $X_8|X_2, X_4$. Frank’s Copula Vine vs Joint Normal Vine (conditioning in Netica using $8 \cdot 10^5$ samples).

A very important feature of this method is that conditioning can be done theoretically. Since all the calculations are performed on a joint normal vine, any conditional distribution will also be a normal with the mean and variance given by the formulas in Proposition 2.2. Finding the conditional distribution of the corresponding original variable will just be a matter of transforming it back using the inverse distribution function of this variable and the standard normal distribution function.

6 Conclusions

Vines and Bayesian belief nets, in combination with copula’s, provide practical and flexible ways of stipulating high dimensional distributions. We have presented the copula-vine approach to continuous BBNs, addressed its computational problems and proposed ways to solve them.

Quite often, real life problems are modelled with large BBNs, which might contain big cycles. In most cases, in the copula-vine approach extra calculations are needed in order to sample a structure. These calculations are numerical evaluations of multiple integrals, which are very time consuming. Moreover, updating such a structure is done by re-sampling the network each time a new policy is evaluated. A large model could run for days.

We have shown how to overcome this limita-

tion by making use of the advantages of BBNs software. We sampled the structure just once, imported the sample file in Netica and performed fast updating for the discretized version of the BBN. Not only do we increase the degree of realism by using continuous variables, but we also considerably decrease the number of quantities that must be assessed for the quantification of the discretized BBN (example 4.1 requires 8 number instead of 12,140).

On the other hand, for a large BBN, which contains nodes with many parents, discretized in fairly many states, a very large sample file is needed in Netica. Furthermore, we presented a very simple example in which unexpected problems appear and make this one time sampling very complicated and time consuming. If this happens for a small BBN, it is very likely that for a larger one, more difficulties will emerge.

These problems are caused by the different ordering of variables in the D-vines that represent the BBN structure. We solved this issue by realizing the rank correlation specification on a joint normal vine. We transformed the rank correlation vine to a partial correlation vine on standard normal variables; computed the correlation matrix and sample from a joint normal distribution with standard normal margins and a given correlation matrix. We transformed back, and in this way, we realized the joint distribution of the initial variables with the specified dependence structure. In the joint normal vine we can re-order the variables and compute the re-ordered partial correlations, using the properties of the joint normal distribution. Hence, no extra calculations are involved. The big advantage is that the computational time reduces from hours to seconds.

The joint normal vine method might seem very similar to the joint normal transform method presented in [10]; [11], but the presence of vines is crucial in avoiding the problems encountered in the later method. In the joint normal transform approach, the rank correlation matrix must be first specified and then induced by transforming distributions to standard normals and generating a dependence structure using the linear properties of the joint normal. Specifying a rank correlation matrix can be a very difficult task. Moreover, it is not always possible to find a product moment correlation matrix generating a given rank correlation matrix via Pearson’s transformation, as showed in chapter 4 from [3]. Using the joint nor-

mal vine approach we avoid this problems because we do not specify a rank correlation matrix, but rather a rank correlation vine.

A very attractive feature of this algorithm is that conditioning can be done theoretically. We are presently working on refining this procedure and plan to publish the results elsewhere.

References

- [1] Kurowicka D and Cooke R.M., *Distribution - Free Continuous Bayesian Belief Nets*, (Proc.Mathematical Methods in Reliability Conference,2004).
- [2] Bedford T.J.,Cooke R.M., *Vines - a New Graphical Model for Dependent Random Variables*, **Ann. of Stat.** 30(4) (2002), 1031-1068.
- [3] Kurowicka D and Cooke R.M., *Uncertainty Analysis with High Dimensional Dependence Modelling* , to be published , 2005.
- [4] Yule, G.U., Kendall, M.G, *An introduction to the theory of statistics*, Charles Griffin & Co. 14th edition, Belmont, California, 1965.
- [5] Pearson, K., *Mathematical contributions to the theory of evolution*, **Biometric**, Series. VI.Series (1907).
- [6] Whittaker, J., *Graphical Models in Applied Multivariate Statistics* , John Wiley and Sons, Chichester, 2005.
- [7] Roelen, A.L.C. and Wever, R. and Hale, A.R and Goossenes, L.H.J. and Cooke, R.M. and Lopuhaa, R. and Simons, M. and Valk, P.J.L., *Casual modeling using bayesian belief nets for integrated safety at airports*, **Risk Decision and Policy**,9 (2004), 207-222.
- [8] Tong, Y.L., *The Multivariate Normal Distribution* , Springer-Verlag, New York, 1990.
- [9] Frank, M.J., *On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$* , **Aequationes Math.**, 19, (1979), 194-226.
- [10] Ghosh, S. and Henderson, S.G, *Properties of the NOTRA Method in Higher dimensions*, Proc of the 2002 Winter Simulation Conference, (2002), 1031-1068.
- [11] Iman, R. and Helton, J., *A comparison of uncertainty and sensitivity analysis techniques for computer models*, tech. report, NUREG/CR-3904 SAND84-1461 RG, Albuquerque, 1985.