Relieving the elicitation burden of Bayesian Belief Networks

B.W. Wisse, S.P. van Gosliga, N.P. van Elst, A.I. Barros
TNO Defence, Security and Safety
P.O. Box 96864, 2509JG The Hague
The Netherlands

Abstract

In this paper we present a new method (EBBN) that aims at reducing the need to elicit formidable amounts of probabilities for Bayesian belief networks, by reducing the number of probabilities that need to be specified in the quantification phase. This method enables the derivation of a variable's conditional probability table (CPT) in the general case that the states of the variable are ordered and the states of each of its parent nodes can be ordered with respect to the influence they exercise. EBBN requires only a limited amount of probability assessments from experts to determine a variable's full CPT and uses piecewise linear interpolation. The number of probabilities to be assessed in this method is linear in the number of conditioning variables. EBBN's performance was compared with the results achieved by applying both the normal copula vine approach from Hanea & Kurowicka (2007), and by using a simple uniform distribution.

1 Introduction

In this paper we consider the case of deriving a discrete conditional probability distribution for a node of a Bayesian belief network based on expert judgement. There are many issues to consider when deriving a conditional probability distribution via expert judgement elicitation. The expert assessors will for example use simplifying heuristics when assessing probabilities to avoid too complex mental reasoning. These heuristics might lead to biased assessments. In addition experts might also be subject to various types of motivational biases. There is the problem of how to select the appropriate experts for the elicitation task and how to properly prepare them for formulating the assessments (e.g. motivating and training them). There is the choice of which method to use to elicit the probabilities: e.g. a probability-scale, probability-wheel, gamble-like or adverb-probability matching method? Renooij (2001) gives a good overview of these issues. Though acknowledging their importance, in this paper we do not consider these issues, but focus on reducing the assessment burden of large discrete conditional probability distributions.

The number of probabilities that need to be specified for a node can grow large very easily. For a node with three states that has a parent node with also three states, 6 probabilities need to be specified to determine its conditional probability table (CPT). An additional second and third parent node with three states would consequently require a table of 18 and 54 probabilities, and so on. Apart from the huge amounts of time it would take to assess all the probabilities for large CPTs, it can also be questioned to what extent assessors can be expected to coherently provide the probabilities at the level of detail required (see e.g. (Miller 1956) on the limitations of human short term memory capacity). The elicitation task thus is considered a major obstacle in the use of BBNs (Druzdzel & Van der Gaag 1995, Jensen 1995).

There are two ways in which the elicitation task for discrete BBNs can be relieved. The first is to make it easier for the assessors to provide the probabilistic assessments required. Van der Gaag, Renooij, Witteman, Aleman & Taal (1999) aim to achieve this by transcribing the conditional probabilities and using a scale containing both numerical and verbal anchors. But the effort needed to assess a full CPT using this method, though reduced, is still exponential in the number of conditioning variables. The second option for relieving the elicitation burden is to reduce the number of probabilistic assessments to be made. This can of course be achieved by reducing the number of conditions (parent nodes) or the number of states of the variables, but such reductions will often be undesirable (e.g. leading to loss of detail needed to inform a decision).

The Noisy-OR model, originally introduced by Kim & Pearl (1983), but more extensively discussed in relation to BBNs by e.g. Heckerman & Breese (1996), reduces the number of probabilities to be specified by making additional assumptions about the underlying causal structure of the variables. For the noisy-OR model, the number of probabilities needed to determine the full CPT is linear in the number of conditioning variables, rather than exponential. Although this can mean a huge reduction in elicitation effort, the assumptions necessary are strong and all the variables in the noisy-OR model need to be binary, which strongly limits the applicability of the method.

The Noisy-MAX model (Díez 1993) can be seen as the extension of the Noisy-OR to multi-valued variables. In this model the CPT is derived from 'marginal conditional' distributions specified for each parent: for each parent the probabilities conditional on this parent node are specified and subsequently the full CPT is derived from these conditional probabilities using the max function. The influences of each of the parent nodes are treated in this model as independent. So the joint influence that the parent nodes exercise is fully determined by their marginal influence and a fixed function. Zagorecki & Druzdzel (2006) have fitted the Noisy-MAX model to suitable nodes from three belief networks for which the CPTs where already specified. The authors found the model to be able to provide a good fit to the CPT in about 50% of the cases they considered.

In this paper we develop and evaluate a methodology, EBBN, for deriving a node's CPT in the general case that the states of the node are ordered and the states of each of its parent nodes can be ordered with respect to the influence these parent nodes have on this node of interest. In this method only a (small) part of the CPT- describing the joint influence of the parents in contrast with the marginal influence elicited in the Noisy-MAX model - is elicited. The conditional probabilities that are not directly elicited are derived using an interpolation method based on the ranks of parent node states. The number of probabilities to be assessed is linear in the number of parent nodes. Since the method approximates the probabilities that are not directly assessed, it will contain inaccuracies. Like Van der Gaag et al. (1999) we therefore propose to regard and use this method as a first step in an iterative procedure of stepwise refinement of probability assessments, like described in (Coupé, Peek, Ottenkamp & Habbema 1999).

While testing this method three relevant alternatives were presented. Bonafede & Giudici (2007) have developed a method for deriving a discrete conditional probability distribution based on the marginal distributions, correlation coefficients and standardised joint moments. Yet, this method also requires all the variables to be binary, and closed-form solutions have only been derived for up to three conditioning variables (parent nodes). Secondly Hanea & Kurowicka (2007) provide a method for determining a CPT based on the copula vine approach (Bedford & Cooke 2002) that uses similar prior information: marginal distributions and adjusted (conditional) rank correlations. This method also provides a means for deriving the CPT in the general case that the variables are ordinal and the influences are monotone, although it is not clear to us if and how the prior assessments needed can be elicited accurately from experts. In Section 4 we compare the results of the method developed in this paper with the copula vine approach of Hanea & Kurowicka, for which the required prior assessments are derived from a fully specified CPT.

Very closely related to our method is the method presented by Tang & McCabe (2007). These authors also propose the use of piecewise linear interpolation to approximate not-elicited conditional probabilities. Furthermore they introduce the concepts of dominant and important factors, whilst we use positive and negative dominance and parent weights. Yet, where Tang & McCabe, like Bonafede & Giudici, restrict their method to work with binary variables only, the method we introduce in this paper works with discrete variables in general, under the above described conditions of ordinality of the variables. It should be noted that the development of our method has taken place independently of that of Tang & McCabe.

In the next section we will introduce our alternative elicitation method for BBNs, EBBN, which is aimed at reducing the elicitation burden. In Section 3 we discuss when we can regard an approximation of a CPT to be 'good', providing the means to assess the performance of the proposed method and compare it with the copula vine approach (Section 4). In the final section we present our conclusions and suggestions for future work.

2 The EBBN Method

We regard the problem of expert assessment of the probability distribution of a discrete variable X_c (a node in a BBN), conditional on a set of two or more discrete variables, which we will denote with $pa(X_c)$ (the set of parent nodes). We require (1) the values of X_c to be ordered, and (2) that the values of each of the elements of $pa(X_c)$ can be ordered such that each of these variables have either a positive or a negative influence on X_c. By stating that $X_k \in pa(X_c)$ has a positive influence on X_c, denoted by $S^+(X_k, X_c)$, we mean that observing a higher value for X_k does not decrease the likelihood of higher values of X_c, regardless of the values of the other variables $pa(X_c) \setminus X_k$. We take assignment $a = \{x_j, \ldots, x_u\}$ to be an instantiation of the set of $pa(X_c) = \{X_j, \ldots, X_u\}$. Formally we define $X_k \in pa(X_c)$ having a positive influence on X_c, $S^+(X_k, X_c)$, as (Wellman 1990): for all values x_c of X_c, for all pairs of distinct values $x_{k,n} > x_{k,o}$ of X_k, and for all possible assignments $a_{\neg k}$ for the set of $pa(X_c) \setminus X_k$,

$$P(X_c > x_c \mid x_{k,n}, a_{\neg k}) \geq P(X_c > x_c \mid x_{k,o}, a_{\neg k}).$$

The definition of a negative influence, $S^-(X_k, X_c)$, is completely analogous and would involve only reversing the above inequality.

We define a conditioning variable $X_k \in pa(X_c)$ to be *positive dominant*, if the following two (sets of) assignments of $pa(X_c)$ lead to the same probabilities: (I) all assignments of $pa(X_c)$ in which X_k is in its most favourable state for high values of X_c and (II) the assignment in which each $X_l \in pa(X_c)$ is in its most favourable state for higher values of X_c (i.e. all $X_p \in pa(X_c)$ with $S^+(X_p, X_c)$ are at their highest value, and all $X_n \in pa(X_c)$ with $S^-(X_n, X_c)$ are at their lowest value).

So if a positive dominant parent is in its most favourable state for high values of X_c, then, regardless of the states of the other parents, X_c will have the same probabilities as when conditional on the assignment in which all parent nodes are in their most favourable state. *Negative dominant* variables are defined analogously.

In the remainder of this section we will first discuss the assessments needed from the expert for the derivation of the CPT of X_c. We will then show how to obtain the CPT from these assessments and end the section with an illustrative example of the method, taken from the Hailfinder network (Abramson, Brown, Edwards, Murphy & Winkler 1996).

2.1 Required assessments

It is assumed that the assessor has confirmed that the values of variable X_c are ordered and that the assessor can order the values of each of the variables $X_k \in pa(X_c)$ such that (s)he judges either $S^+(X_k, X_c)$ or $S^-(X_k, X_c)$ to hold. Then the following assessments are required to determine the CPT for variable X_c with conditioning variables $pa(X_c)$[1]:

1. (ordering). For each of the conditioning variables $X_k \in pa(X_c)$: order the values of X_k such that X_k has either a positive or a negative influence on X_c. Fix and record this ordering of the values and the nature of the influence.

2. (typical probabilities). For each of the values x_c of X_c:

   (a) determine the assignment $pa(X_c) = a_{x_c}$ such that the probability $P(X_c = x_c \mid a_{x_c})$ is as large as possible.

   (b) assess the probabilities $P(X_c \mid a_{x_c})$.

   Due to dominance of one of the conditioning variables $a_{x_{c,min}}$ ($a_{x_{c,max}}$) need not be unique, where $x_{c,min}$ ($x_{c,max}$) is the lowest (highest) value of X_c. Therefore $a_{x_{c,min}}$ ($a_{x_{c,max}}$) is by default set to be the assignment in which all the conditioning variables are in their most favourable state for low (high) values of X_c, referred to as a_{neg} (a_{pos}).

3. (weights). For each of the conditioning variables $X_k \in pa(X_c)$, assess $P(X_c = x_{c,max} \mid a_{neg,k+})$ and $P(X_c = x_{c,min} \mid a_{neg,k+})$, where $x_{c,max}$ and $x_{c,min}$ are resp. the maximum and minimum value of X_c, and $a_{neg,k+}$ is the assignment of $pa(X_c)$ in which X_k is in its most favourable state for high values of X_c, and all $X_l \in pa(X_c) \setminus X_k$ are in their least favourable state for higher values of X_c.

4. (dominance). For each of the conditioning variables $X_k \in pa(X_c)$, determine whether X_k has either no, a positive or a negative dominance over X_c.

2.2 Deriving the CPT

The derivation of the CPT of X_c is done in a two-step procedure, using the assessments from Section 2.1. In the first step we will express the probabilities $P(X_c)$ as a function of an influence factor i. In the second step individual and joint influence factors are determined for all assignments of $pa(X_c)$, which are then used to derive the probabilities $P(X_c)$ from the functions of step 1.

The influence factor i is an expression of the positiveness (or negativeness) of the joint influence of the parent variables $pa(X_c)$ on X_c. It is a function of values of the parent variables, with $0 \leq i(a) \leq 1$. We set $i(a_{neg}) = 0$, where $pa(X_c) = a_{neg}$ is the assignment in which all the conditioning variables are in their most favourable state for low values of X_c (see item 2, Section 2.1). And, at the other extreme, $i(a_{pos})$ is set to 1. For all other assignments $i \in (0, 1)$. If assignment a_2 has a strictly more positive influence on X_c than a_1

[1] As mentioned in Section 1 we will not discuss here how these assessments can be best elicited from the assessor

- i.e. $P(X_c > x_c \mid a_2) > P(X_c > x_c \mid a_1)$ for all x_c - then the influence factor corresponding to a_2 should be bigger than the influence factor corresponding to a_1.

We make use of two separate influence factors: the *individual influence factor* i_k for each conditioning variable $X_k \in pa(X_c)$ and the *joint influence factor* i_{joint}. As will become more clear later on, i_k will contain information about the influences exercised by each of the parent variables individually, i_{joint} about the 'general tendency' of all of the parent influences together.

We determine the individual influence factor i_k for $X_k \in pa(X_c)$ as follows:

$$i_k(x_k) := \begin{cases} \dfrac{\text{rank}(x_k) - 1}{\text{rank}(x_{k,max}) - 1} & \text{if } S^+(X_k, X_c) \\ \dfrac{\text{rank}(x_{k,max}) - \text{rank}(x_k)}{\text{rank}(x_{k,max}) - 1} & \text{if } S^-(X_k, X_c) \end{cases} \tag{1}$$

where the rank of the smallest value is set to be 1 and $x_{k,max}$ is the highest value of X_k. So if $X_k \in \{low,$ $medium, high\}$ has a positive influence on X_c, we find that $i_k(low) = 0$, $i_k(medium) = 0.5$ and $i_k(high) = 1$.

The joint influence factor i_{joint} for assignment $pa(X_c) = a$ is derived as:

$$i_{joint}(a) := \frac{\displaystyle\sum_{\{k: X_k \in pa(X_c)\}} i_k(x_k) \cdot (\text{rank}(x_k) - 1)}{\displaystyle\sum_{\{k: X_k \in pa(X_c)\}} (\text{rank}(x_{k,max}) - 1)} \tag{2}$$

Verify that indeed $i_{joint}(a_{neg}) = 0$ and $i_{joint}(a_{pos}) = 1$. Also note that the individual influence factor of X_k, i_k, is equal to the joint influence factor i_{joint} if $pa(X_c) = \{X_k\}$, i.e. if the set of parents of X_c merely consists of X_k.

Step 1. Estimating $P(X_c)$ as a function of joint influence factor i_{joint}

In this step $P(X_c = x_c)$ is estimated as a function of joint influence factor i_{joint}, for each value x_c of X_c. For this we use the orderings determined at item 1 in Section 2.1, and the assignments a_{x_c} and probabilities $P(X_c = x_c \mid a_{x_c})$ assessed at 2. We construct the piecewise linear functions $f_{x_c} : [0, 1] \to [0, 1]$ through the points $(i_{joint}(a_{x_c}), P(X_c = x_c \mid a_{x_c}))$. It can be easily verified that using these linear interpolations ensures that $\sum_{x_c} f_{x_c}(i) = 1$, i.e. the sum of the probabilities of occurrence of the different values of X_c equals unity for all $i \in [0, 1]$. Coherency requires that if $x_{c,n} > x_{c,m}$, also $i_{joint}(a_{x_{c,n}}) > i_{joint}(a_{x_{c,m}})$. In Figure 1 an example is given for how this estimation of $P(X_c)$ as a function of i_{joint} might look like. In this example $X_c \in \{low, medium, high\}$, and the points $(i_{joint}(a_{x_c}), P(X_c = x_c \mid a_{x_c}))$ are assessed as in Table 1.

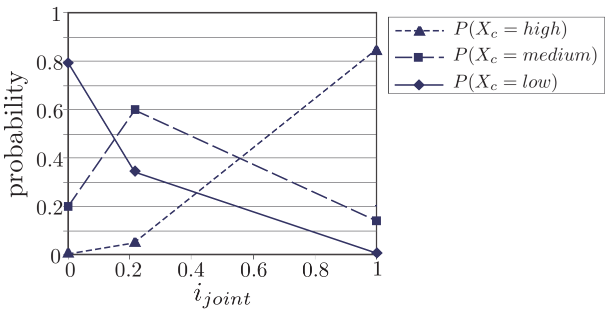

Figure 1: Piecewise linear functions through the points $(i_{joint}(a_{x_c}), P(X_c \mid a_{x_c}))$ from Table 1.

Table 1: Example assessments of $(i_{joint}(a_{x_c}), P(X_c \mid a_{x_c}))$ for $X_c \in \{low, medium, high\}$

x_c	$i_{joint}(a_{x_c})$	$P(X_c \mid a_{x_c})$
low	0	$P(X_c = low \mid a_{low}) = 0.79$ $P(X_c = medium \mid a_{low}) = 0.20$ $P(X_c = high \mid a_{low}) = 0.01$
medium	0.22	$P(X_c = low \mid a_{medium}) = 0.35$ $P(X_c = medium \mid a_{medium}) = 0.60$ $P(X_c = high \mid a_{medium}) = 0.05$
high	1	$P(X_c = low \mid a_{high}) = 0.01$ $P(X_c = medium \mid a_{high}) = 0.14$ $P(X_c = high \mid a_{high}) = 0.85$

Note that $pa(X_c) = a_{low}$ corresponds to the assignment $pa(X_c) = a_{neg}$ and a_{high} to a_{pos}. Hence $i_{joint}(a_{low}) = 0$ and $i_{joint}(a_{high}) = 1$.

Step 2. Deriving the conditional probabilities

In Step 1 we obtained $P(X_c)$ for all possible values of i_{joint} via linear interpolation, and equation (2) provides us with an expression for i_{joint} for all assignments $pa(X_c) = a$. We can now determine $P(X_c \mid a)$ via $P(X_c \mid i_{joint}(a))$ from the functions f_{x_c} of Step 1. Yet this mapping from assignments a for the conditioning variables $pa(X_c)$ to an expression i_{joint} is not unique. Suppose $pa(X_c) = \{X_j, X_k, X_l\}$, X_j and X_l both exercise the same type of influence (positive or negative), and $X_j, X_k, X_l \in \{low, medium, high\}$, then $i_{joint}(\{medium, medium, medium\}) = i_{joint}(\{low, medium, high\}) = 0.5$. As pointed out earlier, i_{joint} is an expression for the 'general tendency' of the influence of the conditioning variables. It does not take into account the (dis)agreement of the influences of each of the conditioning variables individually.

To account for both the 'general tendency' and the individual influences of the conditioning variables, we calculate for each conditioning variable $X_k \in pa(X_c)$ the average of the probabilities $P_k(X_c \mid a)$ over the interval $(min(i_k(x_k), i_{joint}(a)),\ max(i_k(x_k), i_{joint}(a)))$. An example of this average, denoted with $\overline{P_k(X_c \mid a)}$, is illustrated in Figure 2.

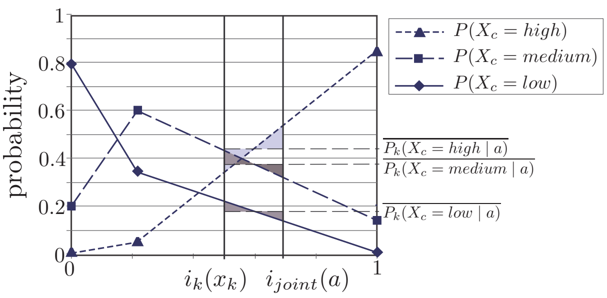

Figure 2: Example of the average probabilities $\overline{P_k(X_c)}$, when $i_k(x_k) < i_{joint}(a)$

We derive the desired probabilities $P(X_c \mid a)$ as the average over the distributions $\overline{P_k(X_c \mid a)}$. Or actually the *weighted* average

$$P(X_c \mid a) = \sum_{k: X_k \mid pa(X_c)} w_k \cdot \overline{P_k(X_c \mid a)}, \qquad (3)$$

since one parent could have a stronger influence on X_c than another. For the same relative change in states, i.e. changes in states resulting in the same absolute change in each of the individual influence factors, the probabilities for X_c might change more for one parent variable than for another. Therefore we calculate the weight w_k for each parent $X_k \in pa(X_c)$, in the following way:

$$w_k = \frac{1}{2} \frac{\delta_k^+}{\sum_{l: X_l \in pa(X_c)} \delta_l^+} + \frac{1}{2} \frac{\delta_k^-}{\sum_{l: X_l \in pa(X_c)} \delta_l^-} \qquad (4)$$

with,

$$\delta_k^+ = P(X_c = x_{c,max} \mid a_{neg,k+}) - P(X_c = x_{c,max} \mid a_{neg})$$
$$\delta_k^- = P(X_c = x_{c,min} \mid a_{neg}) - P(X_c = x_{c,min} \mid a_{neg,k+}).$$

For the derivation of the weights we have taken the situation in which each parent is in its least favourable state for high values of X_c, a_{neg}, as the base. We use the probabilities $P(X_c = x_{c,max} \mid a_{neg,k+})$ and $P(X_c = x_{c,min} \mid a_{neg,k+})$ assessed at item 3 in Section 2.1. Each δ_k^+ and δ_k^- now expresses the changes in the probabilities of resp. the highest and lowest state of X_c, if the one parent X_k is set to its most favourable state for high values of X_c whilst leaving the other parents in their least favourable states ($a_{neg,k+}$). We obtain the weights from these δ's via the normalisations (4).

To a large extent the choice of the base assignment a_{neg} and the probabilities $P(X_c = x_{c,max} \mid a_{neg,k+})$ and $P(X_c = x_{c,min} \mid a_{neg,k+})$ to derive the weights is arbitrary. Even though, we feel the choice for these assignments is one of the most natural choices that can be made. And, more importantly, we feel these assignments are relatively easy for assessors to consider and assess. It is of course possible to use more assessments to determine the weights more accurately. However, we feel that the possible added value does not weigh against the burden of the extra elicitation effort needed.

We derive the desired probabilities $P(X_c \mid pa(X_c) = a)$ by rewriting (3) using (1), (2) and (4), as

$$P(X_c \mid pa(X_c) = a) = \sum_{k: X_k \mid pa(X_c)} w_k \cdot \frac{\int_{i_{min,k}}^{i_{max,k}} \mathbf{f}(i) \cdot di}{i_{max,k} - i_{min,k}} \qquad (5)$$

where $i_{min,k} = min(i_k(x_k), i_{joint}(a))$, $i_{max,k} = max(i_k(x_k), i_{joint}(a))$ and $\mathbf{f}(i) = \left(f_{x_{c,min}}(i), \ldots, f_{x_{c,max}(i)}\right)$.

Finally, we deal with negative and positive dominance of one of the parent variables in the following straightforward way: for all the assignments a_d in which a negative (positive) dominant parent is in its least (most) favourable state for high values of X_c, we set $P(X_c \mid a_d)$ to be equal to $P(X_c \mid a_{neg})$ ($P(X_c \mid a_{pos})$). We will now demonstrate the method by means of an illustrative example.

2.3 Illustrative example from the Hailfinder network

The example given in this section is based on the CompPlFcst variable from the Hailfinder network (Abramson et al. 1996). The variable and its parent nodes, AreaMeso_ALS, CldShadeOth, CldShadeConv and Boundaries, are depicted in Figure 3. For each of the variables also the states (discrete values) are given, ordered and with the highest state on top.

For the variable CompPlFcst we have the fully subjectively specified CPT, consisting of $4 \cdot 3^3 \cdot 3 = 324$ probabilities. In this example we derive the required assessments for EBBN, as specified in Section 2.1, from this CPT, but treat them as if they were directly elicited:

1. (ordering). The ordering of the states of the variables is given in Figure 3, where the highest states are on top. For the conditioning variables we find the following influences:
   S^-(AreaMeso_ALS, CompPlFcst); S^+(CldShadeOth, CompPlFcst); S^-(CldShadeConv, CompPlFcst); S^+(Boundaries, CompPlFcst).

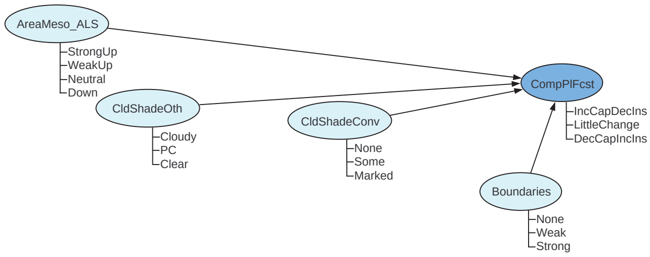

Figure 3: The variable CompPlFcst and its parent nodes from the Hailfinder network.

2. (typical probabilities). We find the assignments: $a_{DecCapIncIns}$ $=\{StrongUp$, $Clear$, $None$, $Strong\}$, $a_{LittleChange}$ $=\{StrongUp$, PC, $Some$, $Strong\}$ and $a_{IncCapDecIns}$ $=\{Down$, $Cloudy$, $Marked$, $None\}$.
   The corresponding conditional probabilities are given in Table 1 and depicted as a function of influence factor i in Figure 1, where $a_{low} = a_{DecCapIncIns}$, $a_{medium} = a_{LittleChange}$ and $a_{high} = a_{IncCapDecIns}$.

3. (weights). As assessments of the remaining probabilities needed to derive the parent weights we find:

$\boldsymbol{a_{neg,k+}}$	$\boldsymbol{P(X_c = x_{c,min} \mid a_{neg,k+})}$	$\boldsymbol{P(X_c = x_{c,max} \mid a_{neg,k+})}$
$a_{neg,\mathsf{AreaMeso\_ALS}+}$	0.20	0.45
$a_{neg,\mathsf{CldShadeOth}+}$	0.40	0.30
$a_{neg,\mathsf{CldShadeConv}+}$	0.52	0.13
$a_{neg,\mathsf{Boundaries}+}$	0.65	0.05

4. (dominance). No (positive or negative) dominant parents.

Now we have all the information (containing only 17 probability assessments!) we need to derive all the 324 probabilities of the CPT of CompPlFcst.

By means of an example we calculate the probabilities $P(\mathsf{CompPlFcst} \mid pa(\mathsf{CompPlFcst}) = a_{expl})$, where a_{expl} $=\{$AreaMeso_ALS$=Down$, CldShadeOth$=PC$, CldShadeConv$=None$, Boundaries$=Strong\}$. For these parent node states we find the individual influence factors: $i_{\mathsf{AreaMeso\_ALS}}(Down) = 1$, $i_{\mathsf{CldShadeOth}}(PC) = \frac{1}{2}$, $i_{\mathsf{CldShadeConv}}(None) = 0$ and $i_{\mathsf{Boundaries}}(Weak) = 0$, and a joint influence factor $i_{joint}(a_{expl}) = \frac{4}{9}$. So we see in this case that the individual influence factors of the parents give a diverse picture, two are very negative (0), one is between negative and positive ($\frac{1}{2}$) and one is very positive (1). This is reflected by the joint influence factor, which has a very average value (0.44), expressing no general tendency of the parent influences towards either positive or negative influence.

Based on the assessments and (4), we find the weights: $w_{\mathsf{AreaMeso\_ALS}} = 0.459$, $w_{\mathsf{CldShadeOth}} = 0.303$, $w_{\mathsf{CldShadeConv}} = 0.165$ and $w_{\mathsf{Boundaries}} = 0.073$. We can now use (5) to derive the desired probabilities and find $P(\mathsf{CompPlFcst} \mid pa(\mathsf{CompPlFcst}) = a_{expl}) = \{0.17, 0.32, 0.51\}$. We can derive the full CPT of X_c (consisting of 324 probabilities) in the same way, requiring in this case only 17 probabilities to be assessed. When we look up the probabilities in de original CPT, we find $P(\mathsf{CompPlFcst} \mid pa(\mathsf{CompPlFcst}) = a_{expl}) = \{0.20, 0.32, 0.48\}$. The probabilities estimated with the methodology are in this case 'not far off'. Yet, before we can assess how well our method approximates the directly assessed probabilities, we first need to discuss how we can measure the quality of the approximation.

3 Approximation of a CPT for a BBN, when is it 'good'?

Assuming you have knowledge of the 'true' probabilities of a certain CPT, how can you assess the quality of an approximation to that CPT? A measure to assess the similarity between two (discrete conditional) probability distributions, with possibly different support, is the Jensen-Shannon divergence (Lin 1991). Based on the Kullback-Leibler divergence, this measure does not take into account the context of the CPT, the belief network. Both Henrion (1989) and Chan & Darwiche (2002) show that inference in a belief network is most sensitive to assessment errors in probabilities that are close to zero or one.

Druzdzel & Van der Gaag (2000) state that, since inaccuracies will influence the output of the belief network, a natural question to ask is how accurate the approximation should be to arrive at satisfacory behaviour of the network. In other words: if the network is constructed to perform specific queries, does the use of approximations still lead to acceptable outcomes on these queries?

Chan & Darwiche (2002) identify three main approaches in the literature to measure the impact of a change in probability in a CPT: measuring the absolute change in the probability of a query, the relative change in the probability of a query or the relative change in the odds of the query, finding the first to be the most prevalent in the literature.

Zagorecki & Druzdzel (2006) give two measures to express the (dis-)similarity of two CPTs for the same conditional distribution: the Euclidian distance and the Kullback-Leibler divergence between the two CPTs. Time and space unfortunately have prohibited us to implement these measures in the current investigation. We have used the following measures to assess the performance of the EBBN in the next section:

m1. Average absolute error in probability.

m2. Average Jensen-Shannon divergence: a measure of the similarity between the 'true' CPT and the approximation to it.

m3. Maximum Jensen-Shannon divergence.

m4. Number of unmatched certainties and impossibilities: the number of times the 'true' and the approximating CPT disagree on probabilities of 0 and 1. As noted above, queries can be very sensitive to extreme probabilities.

m5. % agreement in likelihood ranking: the percentage of scenarios in which the likelihood ranking of the values of the variable is the same for both the 'true' CPT as the approximating CPT. As scenarios all logically possible combinations of values of the neigbouring (i.e. predecessor and descendent) nodes are taken.

m6. % agreement on most likely state: the percentage of scenarios in which the most likely state for the variable is the same for both the 'true' CPT as the approximating CPT. As scenarios all logically possible combinations of values of the neigbouring (i.e. predecessor and descendent) nodes are taken.

4 Performance of EBBN

We have investigated the performance of the methodology by applying it to a well-known belief network from the literature that contained suitable large subjectively assessed CPTs, and comparing its performance with the copula vine approach from Hanea & Kurowicka (2007). We found the Hailfinder network (Abramson et al. 1996) to contain such CPTs.

4.1 Methodology

We have searched for belief networks that contained nodes that satisfy the following requirements:

- the CPT of the node was subjectively assessed,
- the CPT of the node has to be reasonably challenging in size for elicitation from an expert. For this we decided the node needed to have two or more parents, and
- the states of the node are ordered.

We found these networks are difficult to come by. This is not surprising of course, since these networks would require a huge elicitation effort. Practitioners would usually try to avoid having to specify these large CPTs because the elicitation process would be too time consuming, the very problem we are aiming to deal with in this article. In the BBN repository of the University of Pittsburgh[2] we found the Hailfinder network, which does contain 7 nodes that satisfy our requirements.

For the Hailfinder network we created three alternative versions. In each of these alternative versions we replaced the CPTs of the 7 nodes satisfying the above requirements (and kept the remaining CPTs as they were). In the first alternative implementation these CPTs were replaced with the approximations resulting from the method introduced in this paper. We treat the CPTs from the literature as the 'true' CPTs. We assume that the probabilities needed for our methodology would have been assessed as they are in these CPTs and treat the difference between approximations of the method and the corresponding CPTs as inaccuracies of the approximation. So we have *not* tried to find parameters for EBBN that minimise the distance of the resulting CPT to the original, but have derived the parameters needed from the original CPT.

The second alternative implementation has the selected 7 CPTs derived according to the copula vine approach (Hanea, Kurowicka & Cooke 2006). In this approach a normal copula vine is constructed based on the marginal distributions of each variable and its conditioning variables (or actually continuous versions of these discrete marginals) and (conditional) rank correlation coefficients of the variable with each of its conditioning variables. This normal copula vine specifies a joint distribution of the variable and its conditioning variables. Hanea & Kurowicka (2007) describe how the (conditional) rank correlation coefficients can be derived from a CPT. If one was to use the copula vine approach in practice, the marginal distribution of the variable under consideration and the (conditional) rank correlations with each of the conditioning variables would have to be subjectively assessed, which is not a trivial task. Since we are using the copula vine approach as a benchmark here, as a different means of approximating the 'true' CPT, we simply derived this marginal and the correlations from the 'true' CPT. The used marginal and correlations thus represent the best values that could have been obtained in an elicitation process. After construction we took a large sample (we used a sample size of 80,000) from the normal

[2]The belief network models can be found at the network repository of the Decision Systems Laboratory of the Univesity of Pittsburgh (http://genie.sis.pitt.edu/networks.html)

copula vine and estimated the desired copula vine version of the CPT from the frequencies in this sample. We checked that the marginal of the variable under consideration and the marginals of its parents were still as specified for the copula.

Finally we constructed a third alternative implementation of the Hailfinder network in which all altered CPTs consist of uniform distributions for all assignments of conditioning variables, to serve as a second benchmark. We have assessed the performance of our interpolation method and both benchmarks using the measures specified in Section 3. We found the variable InsInMt to be a positive dominant parent of CldShadeConv, and treated it as such in all three alternative implementations.

4.2 Results

The results of the comparison of the 'true' versions of the selected 7 CPTs of the Hailfinder network with each of the three alternative derivations of these CPTs are given in Table 2 [3]. The table displays how the EBBN, copula vine and the uniform versions of the CPTs score on 9 performance measures. The first four measures are the measures m1.-m4. from Section 3 for the direct comparison between the 'true' and the approximating versions of the CPTs. The measures in the last five columns, m1.-m3., m5. and m6. from Section 3, consider posterior probabilities for each of the 7 selected nodes under all possible scenarios for neighbouring nodes, i.e. all logically possible combinations of states of neighbouring nodes (both parent and child nodes).

For the first seven measures in the table we have that the smaller the measure, the better the performance of the approximating CPT on that measure. For the last two columns to opposite holds: the higher the percentage, the better the performance. If a number is underlined in Table 2, this means that the corresponding approximating method (EBBN, copula vine or uniform) has the best performance for that measurement on that variable.

If we look at the underlined values in Table 2, it seems that EBBN and the copula vine versions are of comparable performance on all performance measures apart from 'unmatched 0/1', on which EBBN performs best on all CPTs. It is comforting to see that both EBBN and the copula vine approach clearly perform better than when the CPT is populated with merely uniform distributions. Further investigation reveals that the EBBN method scores relatively well on the so called 'collector' variables CombMoisture, CombVerMo and CombClouds. These are nodes in the Hailfinder network that "summarize information from different sources about moisture, vertical motion and clouds, respectively" (Abramson et al. 1996, p.69). The EBBN method seems a relatively good means to combine similar information from different sources, at least for the Hailfinder network.

[3]The EBBN and copula vine versions of the Hailfinder network (.xdsl format) can be obtained from the authors.

5 Conclusions and discussion

In this paper we have developed a method for deriving large conditional probability tables based on expert judgement, that can hugely reduce the number of assessments needed from the experts. The quantitative assessments needed from the experts are relatively easy to understand: the experts still need to assess only probabilities. We believe that the experts will also be capable of providing the qualitative judgements described in Section 2.1 at items 1, 2(a) and 4.

In order to evaluate the performance of EBBN we applied it to a well-known belief network from the literature, the Hailfinder network. EBBN's performance was compared with the results achieved by applying both the normal copula vine approach from Hanea & Kurowicka (2007), and by using a simple uniform distribution. The results show that EBBN's performance is comparable to the the performance of the normal copula vine approach, and distinctly better than that of the uniform distributions. We believe that the EBBN method can be a valuable tool for subjectively specifying large CPTs.

In the development of the method, the application to a real-life example (the Hailfinder network) has proven very valuable. We would like to test the method on more examples. But, as noted before, because of cost and effort required to elicit large CPTs, these examples are difficult to find. Any help with finding more examples would be greatly appreciated.

It should be noted that the EBBN method does not always lead to a large reduction in the number of probabilities that need to be assessed. In fact, the method could even require more probabilities to be assessed than there are in the CPT. Roughly this occurs when the number of states of the variable for which the CPT is to be derived is greater than the number of conditions (i.e. the number of different assignments of the conditioning variables).

Finally we would like to remark that the interpolation used, in its current form, does not take into account synergetic effects that may exist between conditioning variables.

Acknowledgements

We are grateful to the Institute of Applied Mathematics of the TU Delft for allowing us to use their Unicorn and Uninet software to derive the normal copula vine versions of the 7 CPTs of the Hailfinder network.

References

Abramson, B., Brown, J., Edwards, W., Murphy, A. & Winkler, R. L. (1996), 'Hailfinder: a Bayesian system for forecasting severe weather', *International Journal of Forecasting* **12**, 57–71.

Bedford, T. J. & Cooke, R. M. (2002), 'Vines - a new graphical model for dependent random variables', *Annals of Statistics* **10**(4), 1031–1068.

Bonafede, C. & Giudici, P. (2007), 'Bayesian networks for enterprise risk assessment', *Physica A* **382**, 22–28.

Chan, H. & Darwiche, A. (2002), 'When do numbers really matter?', *Journal of Artificial Intelligence Research* (17), 265–287.

Coupé, V. M., Peek, N., Ottenkamp, J. & Habbema, J. D. F. (1999), 'Using sensitivity analysis for efficient quantification of a belief network', *Artificial Intelligence in Medicine* **17**, 223–247.

Díez, F. (1993), Parameter adjustment in Bayesian networks. the generalized noisy or-gate, *in* 'Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI93)', Morgan Kaufmann, pp. 99–105.

Druzdzel, M. J. & Van der Gaag, L. C. (1995), Elicitation of probabilities for belief networks: combining qualitative and quantitative information, *in* 'Proceedings of the eleventh conference on uncertainty in artificial intelligence', pp. 141–148.

Druzdzel, M. J. & Van der Gaag, L. C. (2000), 'Building probabilistic networks: where do the numbers come from?', *IEEE Transactions on Knowledge and Data Engineering* **12**(4), 481–486.

Hanea, A. & Kurowicka, D. (2007), Mixed non-parametric continuous and discrete Bayesian belief nets, *in* 'Proceedings of the Fifth International Mathematical Methods in Reliability (MMR) Conference', Glasgow.

Hanea, A., Kurowicka, D. & Cooke, R. M. (2006), 'Hybrid method for quantifying and analyzing bayesian belief nets', *Quality and Reliability Engineering International* **22**(6), 613–729.

Heckerman, D. & Breese, J. (1996), 'Causal independence for probabilistic assessment and inference using Bayesian networks', *IEEE Transactions on Systems, Man and Cybernetics* **26**, 826–831.

Henrion, M. (1989), 'Some practical issues in constructing belief networks', *Uncertainty in Artificial Intelligence* **3**, 161–173.

Jensen, A. L. (1995), Quantication experience of a dss for mildew management in winter wheat, *in* M. J. Druzdzel, L. C. Van der Gaag, M. Henrion & F. Jensen, eds, 'Working Notes of the Workshop on Building Probabilistic Networks: Where Do the Numbers Come From?', pp. 23–31.

Kim, J. & Pearl, J. (1983), A computational model for causal and diagnostic reasoning in inference engines, *in* '8th International Joint Conference on Artificial Intelligence', Karlsruhe, West Germany, pp. 190–193.

Lin, J. (1991), 'Divergence measures based on the shannon entropy', *IEEE Trans. Inf. Theory* **37**, 145–151.

Miller, G. (1956), 'The magical number seven, plus or minus two: some limits on our capacity for processing information', *Psychological Review* **63**, 81–97.

Renooij, S. (2001), 'Probability elicitation for belief networks: issues to consider', *The Knowledge Engineering Review* **16:3**, 255–269.

Tang, Z. & McCabe, B. (2007), 'Developing complete conditional probability tables from fractional data for bayesian belief networks', *Journal of Computing in Civil Engineering* **21**(4), 265–276.

Van der Gaag, L. C., Renooij, S., Witteman, C., Aleman, B. & Taal, B. (1999), How to elicit many probabilities, *in* 'Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence 1999', pp. 647–654.

Wellman, M. P. (1990), 'Fundamental concepts of qualitative probabilistic networks', *Artificial Intelligence* **44**(3), 257–303.

Zagorecki, A. & Druzdzel, M. J. (2006), Knowledge engineering for Bayesian networks: How common are Noisy-MAX distributions in practice?, *in* G. Brewka, S. Coradeschi, A. Perini & P. Traverso, eds, 'ECAI 2006, 17th European Conference on Artificial Intelligence.', pp. 482–488.

Table 2: Performance of the three approximation methods on the measures specified in Section 3

	regarding CPT				regarding posteriors in scenarios				
Variable	av abs diff	av Je-Sh	max Je-Sh	unmatched 1/0	av abs diff	av Je-Sh	max Je-Sh	same likelh ranking	same most likely state
EBBN									
CombVerMo	0.074	0.028	0.099	0 (256)	0.034	0.002	0.025	93.6%	96.8%
CombMoisture	0.029	0.010	0.045	3 (64)	0.205	0.031	0.663	72.0%	82.4%
AreaMoDryAir	0.056	0.035	0.148	9 (64)	0.221	0.030	0.178	61.0%	82.0%
CombClouds	0.061	0.021	0.127	0 (27)	0.251	0.025	0.164	78.1%	82.8%
CldShadeOth	0.131	0.058	0.188	18 (144)	0.361	0.047	0.254	60.8%	70.9%
CldShadeConv	0.071	0.034	0.219	1 (36)	0.223	0.029	0.235	62.5%	73.8%
CompPlFcst	0.065	0.013	0.064	0 (324)	0.044	0.003	0.085	91.0%	92.4%
Copula vine									
CombVerMo	0.090	0.053	0.314	76 (256)	0.039	0.002	0.037	92.0%	95.2%
CombMoisture	0.075	0.040	0.153	7 (64)	0.274	0.037	0.607	60.0%	73.6%
AreaMoDryAir	0.053	0.023	0.072	16 (64)	0.195	0.019	0.094	61.0%	84.0%
CombClouds	0.105	0.043	0.133	1 (27)	0.315	0.036	0.240	76.6%	78.1%
CldShadeOth	0.103	0.040	0.127	23 (144)	0.279	0.032	0.265	78.1%	83.3%
CldShadeConv	0.056	0.015	0.067	2 (36)	0.157	0.012	0.079	71.2%	78.8%
CompPlFcst	0.143	0.069	0.408	0 (324)	0.085	0.012	0.428	86.9%	88.3%
Uniform									
CombVerMo	0.219	0.234	0.549	76 (256)	0.120	0.021	0.229	82.4%	82.4%
CombMoisture	0.130	0.117	0.415	7 (64)	0.797	0.205	0.549	23.2%	23.2%
AreaMoDryAir	0.238	0.273	0.520	16 (64)	0.819	0.218	0.524	25.0%	25.0%
CombClouds	0.289	0.199	0.408	1 (27)	0.810	0.188	0.445	28.1%	29.7%
CldShadeOth	0.293	0.225	0.459	23 (144)	0.747	0.175	0.550	26.2%	41.5%
CldShadeConv	0.149	0.092	0.250	2 (36)	0.404	0.072	0.274	31.2%	50.0%
CompPlFcst	0.142	0.063	0.253	0 (324)	0.089	0.011	0.315	83.4%	85.4%

underlined: best score for the three methods.