RC24818 (W0906-119) June 26, 2009
Computer Science

IBM Research Report

Ordering Nodes for Parameter Elicitation in Bayesian Belief Networks

Debarun Bhattacharjya, Lea A. Deleris, Bonnie Ray
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

Ordering nodes for parameter elicitation in Bayesian belief networks

Abstract

Building Bayesian belief networks in the absence of data involves the challenging task of eliciting conditional probabilities from experts. In this paper, we develop analytical methods for determining the order in which parameters are to be elicited, based on a proximity criteria for the distribution of either the entire set of variables, or a subset of variables of primary interest to the analyst. We explore the implications of our results for typical parameter prior distributions used in the learning community, such as the uniform Dirichlet distribution. Through experiments, we compare the influence of the chosen variables of interest on the ordering.

1 INTRODUCTION

Bayesian belief networks [Pearl, 1988; Howard and Matheson, 1984/2005] are graphical models that have become increasingly popular for representing causal relationships and performing inference in probabilistic domains. When there is no available literature to guide construction of the model and no data from which the model can be "learned", it must be built using the domain knowledge of experts. Often the *structure* of a belief network is relatively easy to construct; then the primary task in building these models is the elicitation of the parameters, as characterized by conditional probability distributions [Henrion, 1989; Druzdzel and van der Gaag, 1995; 2000; van der Gaag et al, 1999].

The cognitive biases prevalent during assessments are well known [Kahneman et al, 1982]. However, there are additional challenges to populating belief networks through expert elicitation in practice. For instance, experts may not necessarily report perfectly what they believe due to a lack of interest, lack of familiarity with either process or topic, or fatigue. Also, elicitation can take considerable time and commitment and experts may drop out before completion. This is particularly relevant in the case of elicitation administered in an automated way such as a survey tool. In this paper, we study the following question: for a particular belief network whose structure is known, and where it is assumed that all the parameters for a node in the network will be elicited together, in what order should the parameters be elicited from a single domain expert so as to best achieve a set of identified goals?

If the analyst is able to elicit all the parameters from experts who report exactly what they believe, then the question of finding the "optimal" order for parameter elicitation is moot. However in practice, the question of choosing an order for elicitation is an important one due to reasons stated above. In this introductory research work, we take an analytical approach to addressing the question of optimal parameter ordering for expert elicitation. A primary objective is to identify reasonable criteria for judging the quality of a partially elicited belief network. Using these criteria, a second objective is to understand how the order for elicitation is determined for popular priors used in the learning community. We develop insights for the case when the priors are uninformative and provide guidelines for analysts in the case of other application-specific belief networks. While empirical research on understanding how experts actually react to different orders is also an important topic, similar to the empirical work on understanding how experts actually feel about different probability elicitation tools [Renoij and Wittemen, 1999; Wang and Druzdzel, 2000], we leave that as a topic for future research.

The remainder of the paper is organized as follows. In section 2, we discuss our basic assumptions and notation. In section 3, we describe criteria for selecting the next node to elicit based on proximity to the joint distribution of a set of variables, including analytical results and insights for some particular choices of priors, and a simulation-based algorithm for the general

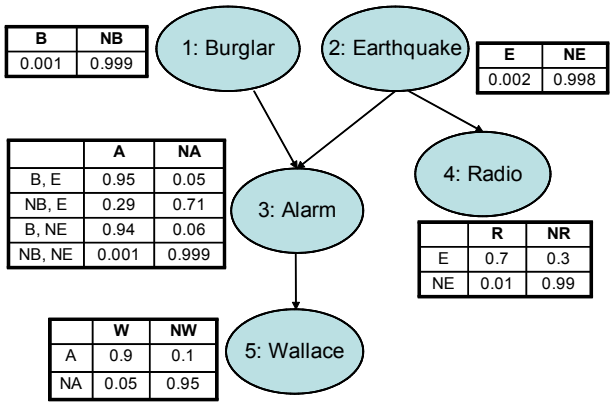

Figure 1: The Burglar network.

problem. Experimental results for some examples are presented in section 4. Finally, section 5 highlights our conclusions and directions for future work.

2 ASSUMPTIONS AND NOTATION

In this section we introduce notation for the paper, and present the assumptions underlying our analytical approach. We assume that the reader is familiar with probabilistic modeling using belief networks. Figure 1 shows a popular belief network from the literature, the Burglar network [Pearl, 1988].

2.1 Preliminaries

We refer to variables and nodes in the belief network interchangeably. We denote variables by uppercase and their instantiations by lower case letters. A set of variables is shown in bold. Let $X_1, X_2, ..., X_N$ be the nodes in the network, where N is the total number of nodes. We will also use $\mathbf{Z}$ to refer to the set of all variables.

We say that x_i is an instantiation (or state) of variable X_i, and $\mathbf{u_i}$ is an instantiation of X_i's parents $\mathbf{U_i}$. A node along with its parents constitutes a family. We denote the conditional probability of x_i given $\mathbf{u_i}$ as $\theta_{x_i|\mathbf{u_i}}$. The probabilities conditioned on $\mathbf{u_i}$ must sum to 1, i.e. $\sum_{x_i} \theta_{x_i|\mathbf{u_i}} = 1$.

We assume that the variables, their states, and the structure of the belief network are known before choosing the order in which the parameters will be elicited. We also assume that all the parameters for a node are elicited together. Hence the phrase "eliciting a node" refers to eliciting all the parameters for that node, which we assume are contained in a conditional probability table (CPT) representation. While some nodes may take longer to elicit due to the number of conditioning parents and their numbers of states,

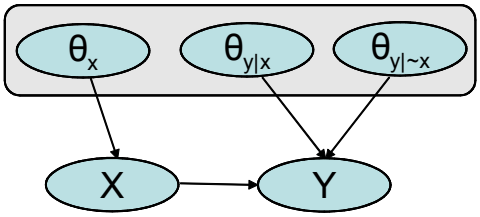

Figure 2: Belief network with nodes for variables X and Y, and their parameters.

we do not consider in this paper the issue of ordering the way in which parameters *within* a node are elicited. From a practical perspective, it seems natural to ask the expert to consider all aspects of a particular scenario at the same time. Indeed, previous research has shown that presenting all conditioning cases for a node together during elicitation reduces the effect of biases [Renooij and Wittemen, 1999; van der Gaag et al, 1999]. When the sizes of the CPT tables are substantially different across nodes, it is of interest to understand the trade-off between the time required to elicit a node relative to the information that can be gained from eliciting the node.

Additionally, we assume that the expert response is exact, i.e. when we elicit a node from an expert, s/he will tell us exactly what s/he believes. While there exists some previous research that studies the effect on elicitation of 'noise' in expert responses, through viewing experts as noisy sensors [Scott and Shachter, 2005], we are not aware of any experimental results that accurately describe how expert responses are affected by node ordering that may, for example, lead to fatigue or generate a learning curve during network elicitation. Thus we focus only on the effect of different node orderings, so as not to confound our results with other effects. The question of the "optimal" order is still interesting even in the case of exact expert response, because experts may drop out of the elicitation process.

2.2 A Bayesian approach

We take a Bayesian approach to uncertainty about parameters in the model, similar to that taken in the learning literature [Heckerman, 1998]. Consider Figure 2, which shows a belief network representing the relationship between two binary uncertainties X and Y and their parameters. Here x and $\sim x$ denote the two states of X. Our uncertainty about the parameters is described using prior distributions, represented by nodes for the θs in Figure 2. For the general problem with nodes $X_1, X_2, ..., X_N$, the parameters are denoted $\theta_{x_i|\mathbf{u_i}}$. We denote their expected values and their standard deviations by $\mu_{x_i|\mathbf{u_i}}$ and $\sigma_{x_i|\mathbf{u_i}}$. Note that pa-

rameters are random variables before they are elicited, but once elicited, they become constants.

We assume (global and local) parameter independence [Spiegelhalter and Lauritzen, 1990], i.e. the parameters for every family are independent of each other, and the parameters for every conditioning case for every node are independent of each other. In Figure 2, these independencies are represented by the lack of arcs between the nodes for the parameters.

A popular prior for parameters is the Dirichlet distribution, since it is conjugate to the multinomial distribution for counts. The Dirichlet distribution for parameter θ with J states is:

$$f(\theta_1, \theta_2, ..., \theta_{J-1}; \alpha_1, \alpha_2, ..., \alpha_J) = \frac{1}{B(\alpha)} \prod_{i=1}^{J} \theta_i^{\alpha_{i-1}}$$

where $\theta_i \geq 0 \forall i$, $\sum_{i=1}^{J-1} \theta_i < 1$, $\theta_J = 1 - \sum_{i=1}^{J-1} \theta_i$ and $B(\alpha)$ is a normalizing constant.

While our methods apply to any prior distribution, we focus on the Dirichlet distribution in this paper. We are particularly interested in the uniform Dirichlet prior, to understand the implications for ordering nodes for parameter elicitation when there is no informative prior information about the parameters.

2.3 A myopic approach

We take a myopic heuristic approach for determining the ordering, i.e. given that a set of nodes (possibly empty) has been elicited, we choose the next node to elicit based only on its effect on the next evaluation of the joint probability distribution, and not on long-term implications of that choice. This approach has also been pursued in the preference elicitation literature [Chajewska et al, 2000; Scott and Shachter, 2005]. In the case of preference elicitation, the graphical model is an influence diagram [Howard and Matheson, 1984/2005], i.e. an extension of the belief network representation that incorporates decisions and values explicitly. In the case of influence diagrams, the value of information [Howard, 1966] is a standard criteria for choosing the next elicitation question. In the next section we discuss criteria for the case of elicitation in a belief network.

3 PROXIMITY-TO-DISTRIBUTION CRITERIA

A belief network represents a joint distribution over a set of variables, based on the expert's beliefs. The joint distribution is typically used for further analysis, such as inference. Moreover, the analyst may be most interested in a specific set of variables in the belief network, which we refer to as the *variables-of-interest*. A possibly incomplete elicitation of a network may result in a distribution that is different from the actual joint distribution, leading to inaccurate inferences. Thus it seems reasonable to require that a possibly incomplete elicitation result in a joint distribution over the variables-of-interest that is "close" to the actual joint distribution over these variables. We discuss appropriate measures of proximity below. We study the case when the variables-of-interest are the entire set of variables $\mathbf{Z}$, followed by a discussion of the case when the variables-of-interest are a subset of $\mathbf{Z}$.

3.1 Proximity to the joint distribution of Z

A natural way to select a node to elicit is to get as "close" to the actual joint distribution of $\mathbf{Z}$ as possible, in the expected sense. Suppose that $P_{\mathbf{Z}}$ is the actual joint distribution of $\mathbf{Z}$, capturing the expert's beliefs about all the uncertain variables, and $Q_{\mathbf{Z}}^{\kappa}$ is the joint distribution of $\mathbf{Z}$ given that only the nodes in the set κ have already been elicited. The joint distribution if node i were to be elicited next ($i \notin \kappa$) is denoted $Q_{\mathbf{Z}}^{i \cup \kappa}$. According to the proximity-to-joint criteria, the next node to elicit is $i^* = argmin_{i \notin \kappa} E\left[D\left(P_{\mathbf{Z}}, Q_{\mathbf{Z}}^{i \cup \kappa}\right)\right]$, for some appropriate distance metric D. After eliciting node i^*, the set κ is updated to include i^*.

Note that in the distribution $Q_{\mathbf{Z}}^{i \cup \kappa}$, the actual parameters for all nodes in the set $\kappa \cup i$ would be known. If the joint distribution were to be used at this stage for inference or some other analysis, all the other parameters would be set to the expected values of the priors, according to our assumptions in Section 2. See Howard (1988) for a discussion on the justification of using the expected values when there is second-order uncertainty.

There are several reasonable distance metrics between probability distributions that can be used for D, such as the Euclidean distance (L_2 norm), the total variation (L_1 norm), or the Kullback-Leibler divergence (which is technically not a distance metric). We focus on the Euclidean distance, for which we can obtain analytical results for the square of this distance (which clearly gives identical results to the Euclidean distance for choosing the next node to elicit). Results for the general problem for any prior and for any distance measure can be obtained through simulation, which we demonstrate later in this section.

The results of Lemma 1, below, provide a dynamic way to choose the next node to be elicited, based on the expert's responses when nodes from the set κ have already been elicited.

Lemma 1. *If nodes from set κ have been elicited, then*

for Euclidean distance squared D, $E\left[D\left(P_{\mathbf{Z}}, Q_{\mathbf{Z}}^{i\cup\kappa}\right)\right] = \sum_{\mathbf{z}}\left[\left(\mu_{x_i|\mathbf{u_i}}{}^2 + \sigma_{x_i|\mathbf{u_i}}{}^2\right)\left(\prod_{j\in\kappa}\theta_{x_j|\mathbf{u_j}}{}^2\right)\left(\prod_{j\neq i;j\notin\kappa}\mu_{x_j|\mathbf{u_j}}{}^2\right)\left(\sum_{j\neq i;j\notin\kappa}\frac{\sigma_{x_j|\mathbf{u_j}}{}^2}{\mu_{x_j|\mathbf{u_j}}{}^2}\right)\right]$ *for node i such that* $i\notin\kappa$*, where* x_i*,* x_j*,* $\mathbf{u_i}$ *and* $\mathbf{u_j}$ *are consistent with an instantiation of all the variables in the network* $\mathbf{z}$*, i.e.* x_i*,* x_j*,* $\mathbf{u_i}$ *and* $\mathbf{u_j}$ *are subsets of* $\mathbf{z}$*.*

Experiments in section 4 explore this result further. We now present a special case of Lemma 1.

Lemma 2. *For Euclidean distance squared D, when all parameters* $\theta_{x_i|\mathbf{u_i}}$ *have the same expected value and variance, and when nodes from set* κ *have been elicited,* $E\left[D\left(P_{\mathbf{Z}}, Q_{\mathbf{Z}}^{i\cup\kappa}\right)\right]$ *is the same for all i.*

When all parameters have the same expected value and variance under the prior distributions, the ordering of nodes for elicitation does not matter, regardless of the expert's responses during elicitation, since eliciting any node yields the same expected distance to the joint at any time during the elicitation. When all nodes have the same number of states, and the priors are uniform Dirichlet, then the condition of equal expected values and variances required in Lemma 2 is satisfied. We now present the result when all priors are uniform Dirichlet, but the nodes may have a different number of states.

Lemma 3. *For Euclidean distance squared D, uniform Dirichlet priors, and when node i has* s_i *states,* $E\left[D\left(P_{\mathbf{Z}}, Q_{\mathbf{Z}}^{i\cup\kappa}\right)\right] = C(\kappa)\left(\frac{s_i}{s_i+1}\right)\left(\left(\sum_{j\notin\kappa}\frac{s_j-1}{s_j+1}\right) - \left(\frac{s_i-1}{s_i+1}\right)\right)$ *where* $C(\kappa)$ *is some function of* κ *and not node i.*

Lemma 3 prescribes a method to choose the next node to elicit, determined only by the number of states for the nodes in the network that have not been elicited yet and not on the number of conditioning parents. Further scrutiny of the expression above shows that the rule for choosing the next node turns out to be a simple, yet non-obvious one. Denote $f(\kappa) = \left(\sum_{j\notin\kappa}\frac{s_j-1}{s_j+1}\right)$. This is a function of the number of states of nodes in the network that have not been elicited yet; this number decreases as every additional node is elicited. If $f(\kappa) > 3$, then the expression in Lemma 3 is monotonically increasing in s_i, so we should select the node with the minimum number of states. However, if $f(\kappa) < 5/3$, then the expression in Lemma 3 is monotonically decreasing in s_i, so we should select the node with the maximum number of states. In the intermediate region, the function first increases then decreases, so we should pick the node with either the minimum or maximum number of states, as determined by Lemma 3.

For example, consider a belief network with 5 nodes, similar to the Burglar network in Figure 1, where nodes 1 to 5 have 3, 4, 5, 6 and 7 states respectively. In this case, by Lemma 3, the ordering of the nodes is: 1, 2, 5, 4, 3. When no nodes have been elicited, $f(\kappa)$ equals 3.23, hence node 1 is chosen because it has the minimum number of states. After node 1 is chosen, $f(\kappa)$ decreases to 2.73 which lies in the intermediate region between 3 and 5/3. Nodes 2 and 5 are chosen in this intermediate region, in that order. Then $f(\kappa)$ equals 1.38, and thus the remaining two nodes elicited are in decreasing order of number of states, nodes 4 and 3.

The intuition behind this seemingly arbitrary rule is as follows. The proximity-to-joint criteria makes a tradeoff between the "spread" in a certain node and the "spread" from a combination of the other nodes not yet elicited. Therefore, whether we should select the node with the minimum or maximum number of states is determined by the degree to which eliciting a node will reduce this spread, which depends, in turn, on the information associated with the nodes that have already been elicited and those yet to be elicited, through the derived expression.

While the assumption of uniform priors provides cases that can be solved easily, solving the general problem in Lemma 1 requires a summation over all instantiations of all variables and is computationally intensive for large networks. However, in practice, a network for a single application may have no relevant prior information, making the assumption of a uniform prior a reasonable one. If not, the node ordering based on Lemma 1 may still be feasible in practice, since networks elicited from experts typically tend to be somewhat small. In the next section, we briefly discuss the case where the variables-of-interest are a (typically small) subset of $\mathbf{Z}$ and show how this framework provides techniques for node ordering that are computationally feasible even for large belief networks.

3.2 Proximity to the joint distribution of Y, a subset of Z

In many belief network applications, the analyst is particularly interested in the joint distribution of a subset of the nodes in the network and not the entire joint distribution. For instance, in networks for medical diagnosis, the analyst is typically interested in the probability that a patient suffers from some disease(s) given the symptom(s). Even though the symptoms will not be known until the patient is observed, the model and the variables (set of symptom variables and disease variables) are known, and the analyst may not be directly concerned about the other variables in the network. A criteria that focuses only on the distribu-

tion of the variables-of-interest seems appropriate.

While it may be argued that a lower number of assessments [Abbas, 2006] can be used in such a situation by eliciting, e.g., marginal distributions directly, often it is most convenient for the expert to assess the distribution for a variable conditioned on other variables [Howard, 1989]. Note that if the analyst is only interested in a subset of variables and there will be no evidence from any other variable in the network, then nodes that are not ancestors of the variables-of-interest can be deleted from the graph based on independencies resulting from d-separation [Pearl, 1988].

To derive our results, we use notation similar to the previous subsection. Suppose that $P_{\mathbf{Y}}$ is the actual joint distribution capturing the expert's beliefs about the variables-of-interest $\mathbf{Y}$, and $Q_{\mathbf{Y}}^{i\cup\kappa}$ is the joint distribution of $\mathbf{Y}$ if node i were to be elicited, given that the nodes in the set κ have already been elicited ($i \notin \kappa$). The next node for elicitation should be $i^* = argmin_{i\notin\kappa} E\left[D\left(P_{\mathbf{Y}}, Q_{\mathbf{Y}}^{i\cup\kappa}\right)\right]$, for some appropriate distance metric D. After eliciting node i^*, the set κ is updated to include i^*.

The following lemma presents a general formulation when D is the Euclidean distance squared.

Lemma 4. *If nodes from set κ have been elicited, then for Euclidean distance squared D, $E\left[D\left(P_{\mathbf{Y}}, Q_{\mathbf{Y}}^{i\cup\kappa}\right)\right] = \sum_{\mathbf{y}}\left[Var[\sum_{\mathbf{z}\sim\mathbf{y}}\{\left(\prod_{j\in\kappa}\theta_{x_j|\mathbf{u_j}}\right)\left(\theta_{x_i|\mathbf{u_i}}\right)\left(\prod_{j\neq i;j\notin\kappa}\theta_{x_j|\mathbf{u_j}} - \prod_{j\neq i;j\notin\kappa}\mu_{x_j|\mathbf{u_j}}\right)\}]\right]$ for node i such that $i \notin \kappa$, where x_i, x_j, $\mathbf{u_i}$ and $\mathbf{u_j}$ in the expression are consistent with $\mathbf{z}$, and $\mathbf{z} \sim \mathbf{y}$ implies that $\mathbf{z}$ is consistent with $\mathbf{y}$.*

While for specific special cases it may be possible to simplify the analytical expression above, the result is not easily interpretable in general. Instead, we now present a general simulation-based algorithm that can be used for any distance metric D and for any prior distribution on the parameters.

3.3 A simulation-based algorithm

The following algorithm uses Monte Carlo simulation to determine the next node to elicit, based on the assumptions in Section 2 and proximity to the joint distribution of the variables-of-interest $\mathbf{Y}$ as the criteria, for any distance metric D. The variables-of-interest can either be all the nodes in the network ($\mathbf{Y} = \mathbf{Z}$), or a strict subset.

Simulation-based Ordering Algorithm: Given a belief network with known variables, states and structure, the variables-of-interest $\mathbf{Y}$, prior distributions on all parameters, and all parameters for CPTs of nodes in set κ (already elicited); determine the next node i^* to elicit ($i^* \notin \kappa$) from the expert as follows:

1. Using the exact values of the parameters for nodes $i \in \kappa$ and a sample point generated from the prior distribution of all parameters for all nodes $i \notin \kappa$, generate a sample point for the actual joint distribution of $\mathbf{Z}$, $P_{\mathbf{Z}}$.
2. Denote by $Q_{\mathbf{Z}}^{i\cup\kappa}$ the partial assessment joint distribution of $\mathbf{Z}$ when node i is elicited, where parameters for node i are the values generated from $P_{\mathbf{Z}}$, parameters for nodes $i \notin \kappa$ are the expected values of the priors, and the parameters for nodes $i \in \kappa$ are the actual numbers already elicited.
3. Find the joint distributions of the variables-of-interest, $P_{\mathbf{Y}}$ and $Q_{\mathbf{Y}}^{i\cup\kappa}\,\forall i$, using inference if required.
4. Find $D\left(P_{\mathbf{Y}}, Q_{\mathbf{Y}}^{i\cup\kappa}\right)\,\forall i$.
5. Repeat Steps 1-3 n times. Use the average value of the n distances to estimate $E\left[D\left(P_{\mathbf{Y}}, Q_{\mathbf{Y}}^{i\cup\kappa}\right)\right]$
6. Based on the obtained average values, choose the next node to elicit $i^* = argmin_{i\notin\kappa}\hat{E}\left[D\left(P_{\mathbf{Y}}, Q_{\mathbf{Y}}^{i\cup\kappa}\right)\right]$ and add i^* to κ.
7. Repeat the entire algorithm to find the next node to elicit.

For some belief networks, there may be a small number of variables-of-interest, maybe even only one variable-of-interest. In such a situation, it is not computationally intensive to find the distributions of $Q_{\mathbf{Y}}^{i\cup\kappa}\,\forall i$, since all these distributions differ from the distribution $Q_{\mathbf{Y}}^{\kappa}$ by parameters associated with only one node. Additionally, recall that $Q_{\mathbf{Y}}^{\kappa}$ is the distribution over the variables $\mathbf{Y}$ when nodes from set κ have been elicited. Previous research has shown that finding the modified probability of evidence when multiple parameters are varied within the same CPT is computationally simple, as it is a special case of n-way sensitivity analysis [Chan and Darwiche, 2004]. Thus the inference queries required in Step 3 of the algorithm can be done efficiently if $\mathbf{Y}$ is a small set. The arithmetic circuit approach to inference [Darwiche, 2003] may be particularly appropriate since it computes partial derivatives with respect to all parameters in a downward sweep. These derivatives can be used for computing the distributions $Q_{\mathbf{Y}}^{i\cup\kappa}\,\forall i$, using the sensitivity analysis method in Chan and Darwiche (2004).

For large belief networks and if $\mathbf{Y}$ is a large set, it may be more efficient to use a function of the marginal

distributions for variables in $\mathbf{Y}$ instead of the joint distribution of $\mathbf{Y}$, since this may be easier to compute. For instance, an additive or multiplicative function of the proximity to marginal distributions in $\mathbf{Y}$ could be used as a criteria for ordering nodes for elicitation. Decomposing a function from a high dimension to its lower dimensional components is a standard approach to simplifying multi-attribute situations. However, this necessarily implies some independence assumptions, which may result in a different ordering from that obtained using the full joint likelihood approach. We leave further exploration of this approach to future research.

4 EXPERIMENTAL RESULTS

4.1 Sensitivity to prior distributions

In this subsection, we investigate the effect of changes in the parameters of the prior distributions on the optimal order. In particular, we consider elicitation of the Burglar network of Figure 1 under different parameterizations of the priors. Initially, we assume that the expected values of the parameters equal their true values (as shown in Figure 1), i.e. we have a "centered" prior, with variances equal to 1/12, corresponding to the variance of a uniform Dirichlet distribution. Our choice of expected values is purely for the purpose of demonstration; in practice, we would of course not know the true values before elicitation. For performing the experiments, we vary the prior distributions systemically over each node, changing the variances *node by node* from 0.001 to 1.0 by increments of 0.001, and changing the expected values *parameter by parameter* from 0.001 to 0.999 by increments of 0.001.

We report here the results obtained for Node 5 in the Burglar network (Wallace), as it is representative of results obtained for other nodes in the network. In our example, Node 5 has two independent parameters: $Pr(Wallace|Alarm) = 0.9$ and $Pr(Wallace|NoAlarm) = 0.05$. Under our base assumptions, the default optimal ordering is: 5, 4, 3, 1, 2. Table 1 shows the ordering of nodes for different ranges of the variance, while Tables 2 and 3 display the node orderings as the expected value of the two independent parameters associated with Node 5 vary from 0 to 1.

We observe that as the variance increases, the importance of the node in the ordering increases as well. Indeed, Node 5 starts in the last position when its variance is below 0.051 and becomes the first node specified for elicitation as its variance increases above 0.077. Additionally, we see that the relative order of the other nodes is not affected. Moreover, the results

Table 1: Sensitivity of optimal ordering to the variance of all parameters associated with Node 5

Variance	1st	2nd	3rd	4th	5th
0.001-0.051	4	3	1	2	5
0.051-0.068	4	3	1	5	2
0.068-0.070	4	3	5	1	2
0.070-0.077	4	5	3	1	2
Above 0.077	5	4	3	1	2

Table 2: Sensitivity of optimal ordering to the expected value of parameter for $(Wallace|Alarm)$

Expected value	1st	2nd	3rd	4th	5th
0.001-0.364	5	4	3	1	2
0.364-0.637	4	5	3	1	2
0.637-1.000	5	4	3	1	2

indicate that the threshold value at which Node 5 gains rank depends on the expected values of the parameters for all nodes (since parameters for all nodes have prior distributions with the same variance). The main observation with respect to Tables 2 and 3 is that the optimal ordering with respect to the expected values of Node 5's parameters is robust, i.e. the optimal order is unchanged if $Pr(Wallace|Alarm) \in [0.637; 1]$ and if $Pr(Wallace|NoAlarm) \in [0.015; 0.986]$. As mentioned above, similar analysis was conducted for each node, with analogous results.

4.2 Effect of the variables of interest

In this section, we explore the effect of the choice of variables-of-interest on the ordering. For this example, we use a belief network adapted from the risk analysis literature [Paté-Cornell and Fischbeck, 1993], the Shuttle network shown in Figure 3. This network was built to evaluate the probability of loss of the space shuttle if debris damaged the thermal protection system of the orbiter. This is the unfortunate event that occurred during the Columbia space shuttle voyage in February, 2003.

In our version of the Shuttle network, there are eight nodes which together decompose accident scenarios beginning with either "Debris Hit" or "Debonding" of the shuttle tiles, that may ultimately lead to "Subsys-

Table 3: Sensitivity of optimal ordering to the expected value of parameter for $(Wallace|NoAlarm)$

Expected value	1st	2nd	3rd	4th	5th
0.001-0.015	4	5	3	1	2
0.015-0.986	5	4	3	1	2
0.986-1.000	4	5	3	1	2

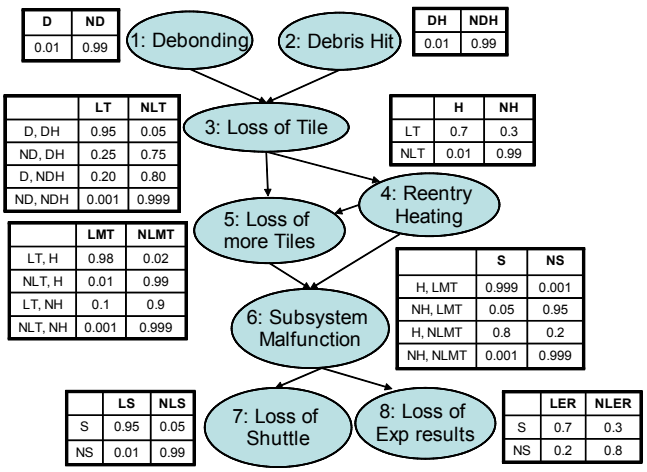

Figure 3: The Shuttle network.

tem Malfunction", which in turn may result in "Loss of Shuttle" and/or "Loss of Experimental Results". Note that "Loss of Shuttle" is a more catastrophic consequence where the shuttle is lost in the accident, whereas "Loss of Experimental Results" implies that the shuttle's records have been erased due to technical malfunctions. We assume that "Loss of Shuttle" and "Loss of Experimental Results" are independent given "Subsystem Malfunction"; it may be possible to retrieve the records even if the shuttle can not return.

Our experiments with this example were performed as follows. For two different sets of priors, we evaluated the optimal ordering for four sets of variables-of-interest:

- the entire joint distribution
- the marginal distribution of "Loss of Shuttle" (Node 7)
- the marginal distribution of "Loss of Experimental Results" (Node 8)
- the joint distribution of "Loss of Shuttle" and "Loss of Experimental Results" (Nodes 7 and 8)

The optimal ordering for the joint is obtained using Lemma 1 and the optimal ordering for the three others are obtained through the implementation of the simulation-based ordering algorithm (100 iterations). The results of our experiments are reported below. Table 4 corresponds to the case where all nodes have uniform Dirichlet priors, and Table 5 to the case where all nodes have Dirichlet priors with the expected values equal to the true values of the parameters ("centered priors") and the variance for all parameters arbitrarily set to $5E-4$.

Table 4: Optimal ordering for four sets of variables-of-interest - uniform Dirichlet priors

All variables	Doesn't matter (Lemma 2)							
Loss of Shuttle	7	6	4	3	2	1	5	8
Loss of Exp. Results	8	6	4	3	2	1	5	7
Loss of Shuttle and Exp. Results	7	8	6	4	3	2	1	5

Table 5: Optimal ordering for four sets of variables-of-interest - centered Dirichlet priors

All variables	8	7	5	6	1	2	4	3
Loss of Shuttle	7	4	1	6	3	2	5	8
Loss of Exp. Results	8	6	4	2	1	5	3	7
Loss of Shuttle and Exp. Results	3	8	4	6	7	2	1	5

The first insight from our experiments is that the choice of variables-of-interest plays a major role in determining the optimal ordering. Therefore, whenever possible, how the belief network is going to be used should be clarified before starting the elicitation process. Second, we observe that for both our assumptions on priors, when focusing on a single node as variable-of-interest, that node comes first in the optimal ordering. This observation is no longer valid when the subset has more than one element: for the case of our "centered priors" with variables of interest being both "Loss of Shuttle" and "Loss of Experimental Results", the first node to be assessed should be Node 3 - "Loss of Tile". For the case of uninformative priors, we notice that the relative order of the other nodes remains similar whether we focus on Node 7, Node 8 or both. This may be the consequence of the fact that all other nodes have the same prior distributions. We do not observe the same behavior when we consider the case of centered priors.

5 CONCLUSION

In this paper, we develop methods for ordering the way in which information is elicited to populate a Bayesian belief network from experts, under different assumptions concerning the prior beliefs as well as different assumptions concerning the variables of primary interest. For specific cases about the prior information, we derive analytical expressions to determine the optimal ordering, and we provide a simulation-based algorithm to determine the ordering in the general case.

These orderings are driven solely by analytical concerns and do not consider the effect on the user-

friendliness of the elicitation process. In practice, belief networks are often elicited in top-down fashion, starting with nodes with no parents and going down the network. One avenue for future research is to investigate the consequence of forcing a possibly unnatural ordering upon experts and to assess whether the "information gain" from an analytical perspective is worth the cost in practice, e.g., confusion, fatigue and increased imprecision.

Acknowledgements

We thank Witi Sachchamarga for many helpful contributions to research discussions.

References

Abbas, A., 2006, Entropy methods for joint distributions in decision analysis, *IEEE Transactions on Engineering Management*, **53(1)**, pp. 146–159.

Chajewska, U., Koller, D., and Parr, R., 2000, Making rational decisions using adaptive utility elicitation, In Chickering, M., and Halpern, J., *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*, pp. 67–75, Banff, Canada: AUAI Press, Alrington, VA.

Chan, H., and Darwiche, A., 2004, Sensitivity analysis in Bayesian networks: From single to multiple parameters, In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pp. 363–369, Austin, Texas: AUAI Press, Alrington, VA.

Darwiche, A., 2003, A differential approach to inference in Bayesian networks, *Journal of the ACM*, **50(3)**, pp. 280–305.

Druzdzel, M., and van der Gaag, L., 1995, Elicitation of probabilities for belief networks: Combining qualitative and quantitative information, In Besnard, P., and Hanks, S., *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 141-148, Montreal, Canada: Morgan Kauffman, San Francisco, CA.

Druzdzel, M., and van der Gaag, L., 2000, Building probabilistic networks: Where do the numbers come from?, Guest editors' introduction, *IEEE Transactions on Knowledge and Data Engineering*, **12(4)**, pp. 481-486.

Heckerman, D., 1998, A tutorial on learning with Bayesian networks, In Jordan, M., editor, *Learning in Graphical Models*, Kluwer, Netherlands.

Henrion, M., 1989, Some practical issues in constructing belief networks, In Henrion, M., Shachter, R., Kanal, L., and Lemmer, J., editors, *Proceedings of the Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 161–173, Windsor, Canada: Elsevier Science, New York, NY.

Howard, R., 1966, Information value theory, *IEEE transactions on Systems Science and Cybernetics*, **2**, pp. 22–26.

Howard, R., 1988, Uncertainty about probability: A decision analysis perspective, *Risk Analysis*, **8(1)**, pp. 91–98.

Howard, R., 1989, Knowledge maps, *Management Science*, **35**, pp. 903–922.

Howard, R., and Matheson, J., 1984/2005, Influence diagrams, In *The Principles and Applications of Decision Analysis* Vol. II, Strategic Decisions Group, Menlo Park, CA, pp. 719–762, Reprinted, *Decision Analysis*, **(3)**, pp. 127–143.

Kahneman, D., Slovic, P., and Tversky, A., 1982, *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Paté-Cornell, M. E., and Fischbeck, P. S., 1993, Probabilistic risk analysis and risk-based priority scale for the tiles of the space shuttle, *Reliability Engineering and System Safety*, **40**, pp. 221–238.

Pearl, J., 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kauffman, San Francisco, CA.

Renooij, S., and Witteman, C., 1999, Talking probabilities: Communicating probabilistic information with words and numbers, *International Journal of Approximate Reasoning*, **22**, pp. 169–194.

Scott, G., and Shachter, R., 2005, Individualizing generic decision models using assessments as evidence, *Journal of Biomedical Informatics*, **38(4)**, pp. 281–297.

Spiegelhalter, D., and Lauritzen, S., 1990, Sequential updating of conditional probabilities on directed graphical structures, *Networks*, **20**, pp. 579–605.

van der Gaag, L., Renooij, S., Witteman, C., Aleman, B, and Taal, B., How to elicit many probabilities, 1999, In Laskey, K., and Prade, H., editors, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 647–665, Stockholm, Sweden: Morgan Kaufmann, San Francisco, CA.

Wang, H., and Druzdzel, M., 2000, User interface tools for navigation in conditional probability tables and elicitation of probabilities in Bayesian networks, In Boutilier, C., and Goldszmidt, M., editors, *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 617–625, Stanford, CA: Morgan Kaufmann, San Francisco, CA.