WP-US-0804

Elicitation Techniques for Bayesian Network Models

Eric Kemp-Benedict

June 6, 2008

Stockholm Environment Institute

Working Paper WP-US-0804

Abstract

This report presents two methods for generating conditional probability tables (CPTs) for Bayesian network models. The methods are based on the notion of departures from a normal state of affairs. The elicitation methods ask experts either for their assessment of the influence of parent nodes on a child node, or for their expectations for the influence of parent nodes on the departure of the child node from a normal state.

For more information about this document,
contact Eric Kemp-Benedict at erickb@sei-us.org.

Stockholm Environment Institute - US
11 Curtis Avenue
Somerville, MA 02144-1224, USA
www.sei-us.org and www.sei.se

Introduction

This report describes two methods for generating full conditional probability tables (CPTs) in Bayesian networks (Pearl, 2001; Ben-Gal, 2007) based on a relatively small number of elicited values from an expert. The goal of such elicited probability tables (EPTs) is to simplify the task for the expert (Cain, 2001; MERIT, 2005) and can also formalize some of the informal thought processes an expert might employ when filling in the full CPT.

It is necessary to elicit probability tables in at least two situations when building Bayesian networks. First, there may be no measured data with which to estimate the CPT, in which case expert elicitation may be the only option (Cain, 2001; MERIT, 2005). Second, there may be insufficient data to definitively establish a CPT. CPTs can grow rapidly in size with the number of parent nodes or states, and if the size of the data set is not considerably larger than the number of parameters, then the CPT cannot be determined from the data alone. In this case, a useful strategy is to elicit an initial CPT that is then modified by training the network with data.

The method for generating EPTs described by Cain (2001) is relatively flexible, but can accommodate only relatively small matrices. The number of elicited parameters, while less than for the full CPT, grows rapidly as the number of nodes or states increases. The methods described in this report remain tractable as the number of nodes and states increases. The strategy in each case is to assume a certain kind of independence between the parent nodes, where the definition of "independence" differs from one method to the other.

The assumption of independence and other assumptions made to simplify the task of elicitation place restrictions on the CPT that might not correspond to reality. This must be borne in mind when these techniques are applied. However, if they are used to generate initial CPTs that are subsequently trained with data, then the limitations on the form of the initial CPT are of less concern. Also, to the extent that the techniques capture the informal thought process that an expert might go through in order to produce parameters for the CPT, formalizing the process can make the exercise simpler. Finally, the generated CPT can be used as a starting point, and then be subsequently reviewed and modified by the expert.

The typical distribution

Both of the methods described in this report takes as a starting point the idea of a "normal" or "typical" state of affairs that is described by a user-defined typical distribution T_r. The specific form of the typical distribution is not important, only that it is used consistently in the network and, conceptually, that it represents in the mind of the expert a normal state of affairs. Examples of possible typical distributions include:

1. Low : Middle : High :: 0.25 : 0.50 : 0.25
2. Very Low : Low : Middle : High : Very High :: 0.05 : 0.20 : 0.50 : 0.20 : 0.05
3. Low : Middle : High :: 0.33 : 0.33 : 0.33

In principle, the typical distribution could also be a continuous distribution, such as a normal distribution with a specified mean and standard deviation. However, it is assumed in the presentation that the typical distribution is discrete.

Method 1: Influence Weights

Consider a conditional probability $P(r|s_1, s_{2,} \ldots, s_N)$ in which all parent and child nodes take one of the *m* possible states of the typical distribution T_r. In this method it is assumed that the CPT has the property that the change in the probability under a change in any parent node s_i is independent of the other parent nodes. That is, if a parent state changes from "low" to "high", then the change in the probability of state *r* is the same no matter what the values for the other parent states.

Background

With the restrictions described above, the CPT can be written

$$P(r|s_1, s_{2,} \ldots, s_N) = \sum_{i=1}^{N} M_{r s_i}^{(i)} , \tag{1}$$

where the *N* matrices $M_{r s_i}^{(i)}$ are each of size $m \times m$. The condition that the conditional probability sum to one then gives

$$\sum_{r=1}^{m} P(r|s_1, s_{2,} \ldots, s_N) = 1 \;\Rightarrow\; \sum_{r=1}^{m} \sum_{i=1}^{N} M_{r s_i}^{(i)} = 1 \; . \tag{2}$$

However, since each entry in this sum is independent of the others, this can only be satisfied if

$$\sum_{r=1}^{m} M_{r s_i}^{(i)} = w_i \; , \tag{3a}$$

$$\sum_{i=1}^{N} w_i = 1 \; . \tag{3b}$$

This is satisfied in general for a matrix of the form

$$M_{r s_i}^{(i)} = w_i \, \delta_{r s_i} + \mu_{r s_i}^{(i)} , \tag{4}$$

where δ_{ij} is equal to one if $i = j$, and equal to zero otherwise.

$$\sum_{r=1}^{N} \mu_{r s_i}^{(i)} = 0 \; . \tag{5}$$

The elicitation process assumes that the matrices $\mu_{r s_i}^{(i)}$ are identically zero. This means that the parent nodes are assumed to be maximally independent from one another. In this case, the elicitation procedure seeks to determine the weights w_i.

Note that when the matrices $\mu_{r s_i}^{(i)}$ are set to zero, then the EPT has the following properties:

1. When all parent nodes are in the same state (e.g., "high"), then the child node is also in that state.
2. When all of the parent nodes have the typical distribution T_r, then the child node also has the typical distribution.

In fact, the second property follows from the first property. The first property is rarely encountered in practice, since there will usually be influences other than those of the identified parent nodes. To capture this, the elicitation method adds an "environmental" parent node that is not explicitly present in the network. This parent node always has the typical distribution T_r. In this case, the first property changes to

1. When all parent nodes are in the same state, then the probability that the child node is in that state is at its maximum,

while the second property still holds.

Elicitation procedure

Combining Equations (1) and (4), setting $\mu_{r s_i}^{(i)}$ to zero, and adding the environmental node, the EPT described above can be written

$$P(r|s_1, s_2, \ldots, s_N) = \sum_{i=1}^{N} w_i \delta_{r s_i} + \left(1 - \sum_{i=1}^{N} w_i\right) T_r \,. \tag{6}$$

The sum of the weights is referred to as the "information" provided by the parent nodes, and the elicitation proceeds in three steps:

1. Identify the most influential parent node (or nodes). Give this node (or set of equally influential nodes) a value for a "non-normalized" weight $\hat{w}_i$ of 10.
2. Assign relative non-normalized weights to all other nodes. For example, one that is estimated to be one-half as influential as the most influential node would receive a weight of 5.
3. Estimate the amount of information provided by the parent nodes as a percentage. This can be asked in the following form: "Knowing the values of the parent nodes, how certain would you be about the value of the child node?" This is the information *I*.

With this information, the normalized weights are calculated as

$$w_i = I \hat{w}_i \left(\sum_{i=1}^{N} \hat{w}_i \right)^{-1} \,. \tag{7}$$

Because the the elicitation procedure involves the specification of weights to each of the parent nodes, this is called the "influence weight" method.

Method 2: Likelihood

The method of influence weights requires that each of the parent nodes has the same states as the typical distribution. This is not always realistic. Also, it assumes a particular form of independence between the parent nodes – that a change in state in any child node always produces the same change in the CPT – and this assumption may not match the expectation of the expert. In this case, this second method may be applicable. It does not assume that the parent nodes have the same states as the typical distribution, although the child node does, and the independence assumption is applied to a log likelihood, rather than to the CPT itself.

Background

This method takes Bayes' rule (Bayes, 1763; Gill, 2002) as its starting point. This can be written,

$$P(r|s_1, s_2, \ldots, s_N) \propto L(r|s_1, s_2, \ldots, s_N) P(r) \,, \tag{8}$$

where $P(r)$ is the prior probability for r and $L(r|s_1, s_2, \ldots, s_N)$ is the likelihood of r given the states for each of the parent nodes. In this method the prior probability for r is given by the typical distribution, so that

$$P(r|s_1, s_2, \ldots, s_N) \propto L(r|s_1, s_2, \ldots, s_N) T_r . \tag{9}$$

The interpretation in this case is that in the absence of any information about the parent nodes, it can be assumed that the child node has the typical distribution. Otherwise, the parent nodes tend to move the probability of the child node away from the typical distribution in systematic ways. This is determined by the likelihood, and so this method is called the "likelihood" method.

It is convenient (and conventional) to use the log likelihood, rather than the likelihood itself, because log likelihood typically covers a smaller range of values. The key assumption in this method is that log likelihood can be expressed as a sum of independent terms, one for each of the parent nodes. That is,

$$\log_b L(r|s_1, s_2, \ldots, s_N) = \sum_{i=1}^{N} a_{r s_i}^{(i)} . \tag{10}$$

When one of the parameters $a_{r s_i}^{(i)}$ is equal to zero, then it contributes a factor of $b^0 = 1$ to the likelihood, meaning that it does not shift the probability distribution of the child node away from the typical distribution. If it is positive, then it increases the likelihood for that combination of r and s_i, and if it is negative then it makes it less likely. The base b for the logarithm can be chosen so that the values for the parameters $a_{r s_i}^{(i)}$ are of a convenient magnitude.

Elicitation procedure

The elicitation procedure amounts to eliciting the parameters $a_{r s_i}^{(i)}$ and then calculating the CPT using Equations (9) and (10). One convenient variant on this method is to define a set of numbers x_r such that

$$a_{r s_i}^{(i)} = x_r \alpha_{s_i}^{(i)} . \tag{11}$$

For example, x_r might take values:

1. High : Medium: Low :: 1 : 0 : -1
2. Very high: High: Medium : Low : Very Low :: 1.0 : 0.5 : 0.0 : -0.5 : -1.0

In this way, only one value must be specified for each state of each parent node (the parameters $\alpha_{s_i}^{(i)}$).

When to use each method

The first method (influence weights) is indicated if the parent nodes to a child node has the same states as the typical distribution, and the mental model of the expert satisfies:

1. Typical inputs should give a typical output,
2. If all inputs are all in the same state, then the the probability of the child node being in that state is at its maximum.

In fact, this method can be used even if some of the parent nodes do not have the states of the typical distribution, as long as the states act as labels for different options or situations. Such nodes are what Cain (2001) calls "modifying" nodes. For example, a Bayesian network might be applied to several agro-ecological zones (AEZs). In this case, there may be an AEZ node whose states are the agro-ecological zones, while each of the other parent nodes reflects farmer choices, climate, or other variables whose states match those of the typical distribution. In this case, a different CPT can be elicited for each of the AEZs.

The second method (likelihood) is called for if the mental model of the expert is that of parent nodes driving the child node away from a typical distribution. For example, a parent node may be rainfall, and the child node may be yield. Low rainfall would shift the probability distribution for the yield toward lower values.

In either case, the use of EPTs is a time-saving device that may lead to unrealistic results. It is important for the expert to explore the behavior of the EPT within the network to determine whether it is behaving sensibly, according to his or her judgment.

References

Bayes, T. 1763. "An essay towards solving a problem in the doctrine of chances, by the late Rev. Mr. Bayes, Communicated by Mr. Price, in a letter to John Canton, M.A. and F.R.S.", *Philosophical Transactions of the Royal Society of London* **53**:370-418.

Ben-Gal, I. 2007. "Bayesian networks". In F. Ruggeri, R. Kenett, and F. Faltin, editors, *Encyclopedia of statistics in quality and reliability*, John Wiley & Sons.

Cain, J. 2001. *Planning Improvements in Natural Resources Management.* Wallingford, UK: Centre for Ecology and Hydrology.

Gill, J. 2002. *Bayesian methods : A social and behavioral sciences approach*. Boca Raton, Fla.: Chapman & Hall/CRC Press.

Management of the Environment and Resources using Integrated Techniques (MERIT). 2005. *Guidelines for the use of Bayesian networks as a participatory tool for water resource management: based on the results of the MERIT project*. Compiled by John Bromley. Wallingford, UK: Centre for Ecology and Hydrology.

Pearl, J. 2001. "Causal inference in statistics: a gentle introduction". In *Computing science and statistics, proceedings of Interface '01 33*. UCLA CSL Technical Report (R-289), August 2001. [online] URL: ftp://ftp.cs.ucla.edu/pub/stat_ser/R289.pdf.