Technical Report No. 5 April 18, 2014

Bayesian Networks

Michal Horný
mhorny@bu.edu

Boston University School of Public Health
Department of Health Policy & Management

This paper was published in fulfillment of the requirements for PM931: Directed Study in Health Policy and Management under Professor Cindy Christiansen's (cindylc@bu.edu) direction. Jake Morgan, Marina Soley Bori, Meng-Yun Lin, and Kyung Min Lee provided helpful reviews and comments.

Contents

Executive summary

A Bayesian network is a representation of a joint probability distribution of a set of random variables with a possible mutual causal relationship. The network consists of *nodes* representing the random variables, *edges* between pairs of nodes representing the causal relationship of these nodes, and a *conditional probability distribution* in each of the nodes. The main objective of the method is to model the posterior conditional probability distribution of outcome (often causal) variable(s) after observing new evidence. Bayesian networks may be constructed either manually with knowledge of the underlying domain, or automatically from a large dataset by appropriate software.

KEYWORDS: Bayesian network, Causality, Complexity, Directed acyclic graph, Evidence, Factor, Graphical model, Node.

1 Introduction

Sometimes we need to calculate probability of an uncertain cause given some observed evidence. For example, we would like to know the probability of a specific disease when we observe symptoms in a patient. Such problems are often notably complex with many inter-related variables. There might by many symptoms, and even more potential causes. In practice, it is usually possible to obtain only the reversed conditional probability, i.e. probability of the evidence given the cause, the probability of observing symptoms if the patient has the disease. A Bayesian approach is appropriate in these cases, while *Bayesian networks*, or alternatively *graphical models*, are very useful tools for dealing not only with uncertainty, but also with complexity and (even more importantly) causality, Murphy (1998). Bayesian networks have already found their application in health outcomes research and in medical decision analysis, but modelling of causal random events and their probability distributions may be equally helpful in health economics or in public health research.

This technical report presents a brief overview of the method and provides the reader with basic instructions to apply the method in research practice. Section 2 briefly explains the theoretical background of Bayesian networks. Subsequently, section 3 presents instructions on how to build a Bayesian network. Section 4 overviews available software and finally section 5 can be used as a guide through helpful literature for further study.

2 Theoretical background

A Bayesian network represents the causal probabilistic relationship among a set of random variables, their conditional dependences, and it provides a compact **representation of a joint probability distribution**, Murphy (1998). It consists of two major parts: a **directed acyclic graph** and a set of **conditional probability distributions**. The directed acyclic graph is a set of random variables represented by nodes. For health measurement, a node may be a health domain, and the states of the node would be the possible responses to that domain. If there exists a causal probabilistic dependence between two random variables in the graph, the corresponding two nodes are connected by a directed edge, Murphy (1998), while the directed edge from a node A to a node B indicates that the random variable A causes the random variable B. Since the directed edges represent a **static** causal probabilistic dependence, cycles are not allowed in the graph. A conditional probability distribution is defined for each node in the graph. In other words, the conditional probability distribution of a node (random variable) is defined for every possible outcome of the preceding causal node(s).

For illustration, consider the following trivial example from Cowel et al. (1999). Suppose we attempt to turn on our computer, but the computer does not start (observation/evidence). We would like to know which of the possible causes of computer failure is more likely. In this simplified illustration, we assume only two possible causes of this misfortune: *electricity failure* and *computer malfunction*. The corresponding directed acyclic graph is depicted in

Figure 1: Directed acyclic graph representing two independent possible causes of a computer failure.

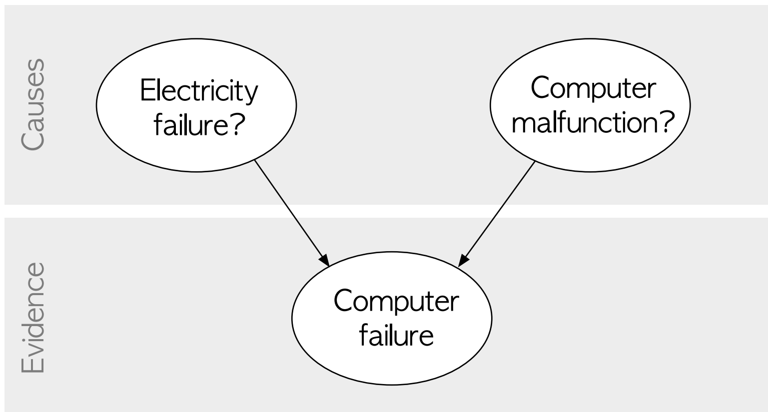


figure 1. The two causes in this banal example are assumed to be independent (there is no edge between the two causal nodes), but this assumption is not necessary in general. Unless there is a cycle in the graph, Bayesian networks are able to capture as many causal relations as it is necessary to credibly describe the real-life situation.

Since a directed acyclic graph represents a hierarchical arrangement, it is unequivocal to use terms such as *parent*, *child*, *ancestor*, or *descendant* for certain nodes, Spiegelhalter (1998). In figure 1, both *electricity failure* and *computer malfunction* are ancestors and parents of *computer failure*; analogically *computer failure* is a descendant and a child of both *electricity failure* and *computer malfunction.*

The **goal** is to calculate the posterior conditional probability distribution of each of the possible **unobserved causes** given the **observed evidence**, i.e. $\mathrm{P}\left[\mathrm{Cause} \,|\, \mathrm{Evidence}\right]$. However, in practice we are often able to obtain only the converse conditional probability distribution of observing evidence given the cause, $\mathrm{P}\left[\mathrm{Evidence} \,|\, \mathrm{Cause}\right]$. The whole concept of Bayesian networks is built on Bayes theorem, which helps us to express the conditional probability distribution of cause given the observed evidence using the converse conditional probability of observing evidence given the cause:

$$\mathrm{P}\left[\mathrm{Cause} \,|\, \mathrm{Evidence}\right] \quad = \quad \mathrm{P}\left[\mathrm{Evidence} \,|\, \mathrm{Cause}\right] \cdot \frac{\mathrm{P}\left[\mathrm{Cause}\right]}{\mathrm{P}\left[\mathrm{Evidence}\right]}$$

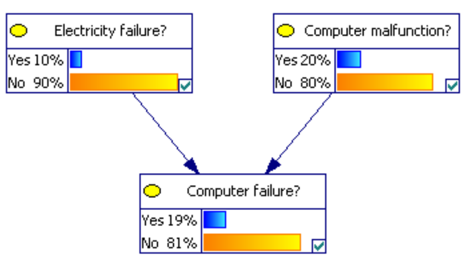
Any node in a Bayesian network is always **conditionally independent** of its all non-descendants given that node's parents. Hence, the joint probability distribution of all random variables in the graph **factorizes** into a series of conditional probability distributions of random variables given their parents. Therefore, we can build a full probability model by only specifying the conditional probability distribution in every node, Spiegelhalter (1998).

Getting back to our example, we suppose that electricity failure, denoted by E, occurs with probability 0.1, $\mathrm{P}[E = \text{yes}] = 0.1$, and computer malfunction, denoted by M, occurs with probability 0.2, $\mathrm{P}[M = \text{yes}] = 0.2$. It is reasonable to assume electricity failure and computer malfunction as independent. Furthermore we assume if there is no problem with the electricity and the computer has no malfunction, the computer works fine. In other words, if C denotes the computer failure, then $\mathrm{P}[C = \text{yes} \mid E = \text{no},\ M = \text{no}] = 0$. If there is no problem with electricity, but the computer has a malfunction, the probability of computer failure is 0.5, $\mathrm{P}[C = \text{yes} \mid E = \text{no},\ M = \text{yes}] = 0.5$. Finally, if the electricity is shut down, the computer will not start regardless its potential malfunction, $\mathrm{P}[C = \text{yes} \mid E = \text{yes},\ M = \text{no}] = 1$ and $\mathrm{P}[C = \text{yes} \mid E = \text{yes},\ M = \text{yes}] = 1$. In this setting, the probability of computer failure $\mathrm{P}[C = \text{yes}]$ can be calculated as

$$\begin{aligned}
\mathrm{P}[C = \text{yes}] &= \sum_{E,M} \mathrm{P}[C = \text{yes},\ E,\ M] \\
&= \sum_{E,M} \Big(\mathrm{P}[C = \text{yes} \mid E,\ M] \cdot \mathrm{P}[E] \cdot \mathrm{P}[M] \Big) \\
&= 0.19
\end{aligned}$$

We can understand the probability $\mathrm{P}[C = \text{yes}] = 0.19$ as a prior (general) probability of computer failure, before we observe any evidence. The graphical model with **prior probability distribution**, i.e. before observing any evidence, is depicted in figure 2.

Figure 2: Directed graphical model representing two independent potential causes of computer failure with prior probability distribution, i.e. before observing any evidence.

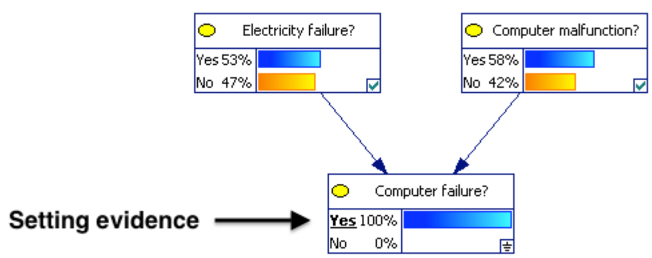


Assume now that we had attempted to turn the computer on, but it did not start. In other words, we **observe** $C = \text{no}$ with probability 1 and we wonder how the probability distribution of electricity failure E and computer malfunction M changed given the observed evidence. Using the Bayes formula, we find

$$
\begin{aligned}
\mathrm{P}\left[E=\text{yes}\,|\,C=\text{yes}\right] &= \sum_M \mathrm{P}\left[E=\text{yes},\ M\,|\,C=\text{yes}\right] \\
&= \sum_M \frac{\mathrm{P}\left[C=\text{yes}\,|\,E=\text{yes},\ M\right]\cdot\mathrm{P}\left[E=\text{yes}\right]\cdot\mathrm{P}\left[M\right]}{\mathrm{P}\left[C=\text{yes}\right]} = 0.53 \\
\mathrm{P}\left[M=\text{yes}\,|\,C=\text{yes}\right] &= \sum_E \mathrm{P}\left[E,\ M=\text{yes}\,|\,C=\text{yes}\right] \\
&= \sum_E \frac{\mathrm{P}\left[C=\text{yes}\,|\,E,\ M=\text{yes}\right]\cdot\mathrm{P}\left[E\right]\cdot\mathrm{P}\left[M=\text{yes}\right]}{\mathrm{P}\left[C=\text{yes}\right]} = 0.58
\end{aligned}
$$

The graphical model with **posterior probability distribution**, i.e. after observing evidence (computer failure), is depicted in figure 3.

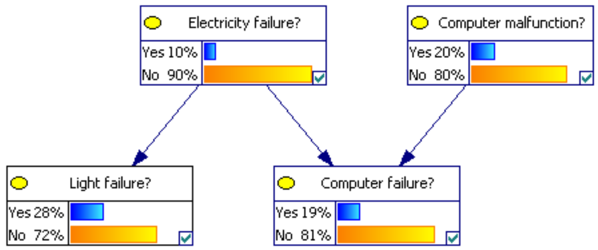
Figure 3: Directed graphical model representing two independent potential causes of computer failure with posterior probability distribution, after observing evidence.



Note that the observed failure has induced a strong dependency between the originally independent possible causes; for example, if one cause could be ruled out, the other must have occurred, Cowel et al. (1999). Nevertheless, the above results are still not very helpful. Assume an extension of the example by incorporating **another piece of evidence** in the model, specifically a light failure L. We assume that light failure is independent of computer malfunction. As before, if the electricity is shut off, the light will not shine under any circumstances, $\mathrm{P}\left[L=\text{yes}\,|\,E=\text{yes}\right]=1$. If there is no problem with the electricity, we assume still a 0.2 chance that the light will go off (broken light-bulb), $\mathrm{P}\left[L=\text{yes}\,|\,E=\text{no}\right]=0.2$. Using the same algorithm as before, we obtain that the prior probability $\mathrm{P}\left[L=\text{yes}\right]=0.28$. The extended graphical model with **prior probability distribution**, before observing any evidence, is depicted in figure 4.

Figure 5 shows changes in posterior probability distribution after observing evidence for all four combinations of light failure and computer failure outcomes. For example, if we

Figure 4: Directed graphical model representing two independent potential causes of computer failure a one potential cause of light failure with prior probability distribution, i.e. before observing any evidence.



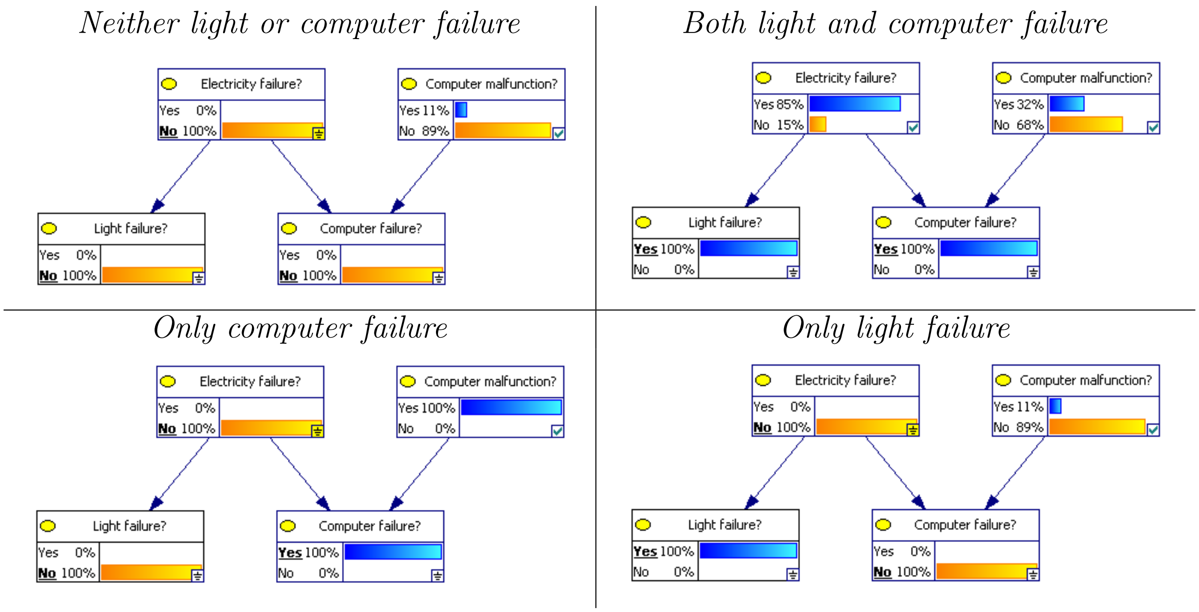
observe both computer failure and light failure, i.e. we observe both $C = \text{yes}$ and $L = \text{yes}$ with probability 1 (top right graph in figure 5), we obtain $\mathrm{P}\left[E = \text{yes} \,|\, C = \text{yes},\ L = \text{yes}\right] = 0.85$ and $\mathrm{P}\left[M = \text{yes} \,|\, C = \text{yes},\ L = \text{yes}\right] = 0.33$. Observation that both the light and computer do not work has substantially increased the chance of electricity failure (there is still a little chance that the light-bulb is broken and the computer has a malfunction). The original computer fault has thus been explained away. In the remaining three cases, at least one of the appliances (light and computer) works, and therefore we may claim that there is nothing wrong with the electricity for sure. If the light works, but the computer does not start (the lower left graph in figure 5), we know for sure that there is nothing wrong with the electricity, therefore computer malfunction is the only possible explanation of the computer failure.

In practice, Bayesian networks are substantially more complex than our example, using tens or even hundreds of nodes. It is also important to note that every node in a graph should be connected with at least one edge to another node. Otherwise, the separated node is independent to all remaining nodes (also to the outcome variable), and therefore there is no need to take this node into account.

Thanks to their visual appearance, Bayesian networks may be confused with Markov models. However, there is a fundamental difference between these two concepts. A Markov model is an example of a graph which represents only **one** random variable and the nodes represent possible realizations of the random variable in distinct time points. In contrast, each node in a Bayesian network represents one random variable in an instant. In other words, Markov models capture dynamics of a single random variable, while Bayesian networks capture static causal relationship among a set of random variables.

Bayesian networks are particularly strong in their ability to **capture causality** and by their **intuitively appealing interface**, Murphy (1998), which helps to effective communication between statisticians and non-statisticians (e.g. physicians or policy-makers), Airoldi (2007). Furthermore, Bayesian networks can be used for **both qualitative and quantitative** modelling, Cowel et al. (1999), since they can **combine objective empirical**

Figure 5: Directed graphical model representing two independent potential causes of computer failure a one potential cause of light failure with posterior probability distribution, i.e. after observing evidence.



probabilities (frequencies) with **subjective estimates**. An important practical strength of Bayesian networks is that they **can be constructed automatically from databases** (so called "learning"), Murphy (1998). Finally, Bayesian networks are able to deal with issues like data over-dispersion (by adding another node representing an additional error term to mean of every observation), relationship between coefficients (representing the coefficients as nodes in the graph), missing data (each missing observation is represented as a node in the graph), measurement errors on covariates, measurement errors on observables, or further sources of complexity. For more details see Spiegelhalter (1998).

Bayesian networks can also be used as **influence diagrams** instead of decision trees. Compared to decision trees, Bayesian networks are usually more compact, easier to build, and easier to modify. Unlike decision trees, Bayesian networks may use direct probabilities (prevalence, sensitivity, specificity, etc.). Each parameter appears only once in a Bayesian network and in case of need, the network may transform into a decision tree, while the reverse is not always possible.

The main weakness is that Bayesian networks **require prior probability distributions**; and despite innocuous choices, these can have misleading effects on the results, Spiegelhalter (1998). Moreover, need for a fully parametrized probability model generally rules out the use of procedures that, although not optimal for specific model assumptions, are robust to a wide range of true situations, Spiegelhalter (1998).

3 How to build a Bayesian network?

There are two ways to build a Bayesian network: a manual construction or automatic construction (so called "learning") from databases. Both methods have advantages and disadvantages.

3.1 Manual construction

Manual construction of a Bayesian network assumes prior expert knowledge of the underlying domain. The first step is to build a directed acyclic graph, followed by the second step to assess the conditional probability distribution in each node.

Directed acyclic graph: Building the directed acyclic graph starts with identification of relevant nodes (random variables) and structural dependence among them, Cowel et al. (1999), Lucas et al. (2004), Airoldi (2007). Not all variables have to be observed; actually some random variables may specify unobserved quantities that are believed to influence the observable outcomes. Data, latent variables and parameters are all considered uniformly as nodes in the graph. However, the underlying conditional probability distribution needs to be known, or at least assumed (e.g. normal distribution). The Bayesian approach is based on assuming all unknown quantities to be random variables, and hence it is natural to include parameters as nodes in a graph, as well as all latent variables and potentially observable quantities. The next step is to sketch the network, Airoldi (2007), taking relationships among the random variable into account, Lucas et al. (2004). The graph structure is usually based on substantive knowledge, although model criticism and revision are often essential, Spiegelhalter (1998).

Despite their name, Bayesian networks do not necessarily imply influence by Bayesian statistics, Murphy (1998). Indeed, it is common to use frequentists' methods to estimate the parameters of the conditional probability distribution. Of course it is possible to implement Bayesian approach by using hyper-parameters instead, Airoldi (2007), i.e. the parameters of the conditional probability distributions underlying the graph could themselves be considered as nodes in the model.

Conditional probability distribution: The constructed directed acyclic graph has to include conditional probability distributions for every node in the graph, Lucas et al. (2004). If the variables are discrete, this can be represented as a table (multinomial distribution), which lists the probability that the child node takes on each of its different values for each combination of values of its parents. If the conditional probability distribution is not available, other statistical methods may be applied to derive this conditional distribution from data (e.g. empirical conditional probability distribution/frequencies estimation). Possible computational methods are outlined e.g. in Spiegelhalter (1998), or Lucas et al. (2004). At this point, the Bayesian network is fully specified. However, it is necessary to perform a sensitivity analysis before the network can be used in real-life application, Lucas et al. (2004). The sensitivity analysis may be performed either as one-way deterministic sensitivity analysis (i.e. varying one parameter at a time over a specified range), or as a probabilistic sensitivity analysis (i.e. varying all parameters of the network at once over a specified probability distribution).

3.2 Automatic learning

Unlike manual construction, automatic learning does not require expert knowledge of the underlying domain. Bayesian networks may be learnt automatically straight from databases using experience-based algorithms often built-in in appropriate software. However, the disadvantage is that automatic construction puts more requirements on the data. Most automatic learning algorithms require no missing data in the dataset, which is often a very strong assumption in practice. If there are missing data in the dataset, these have to be imported, imputated or estimated from other sources, Lucas et al. (2004). Also, there has to be enough data to satisfy the algorithm's requirements for reliable estimates of the conditional probability distributions. For manual construction, the conditional probability distributions are assumed to be a priori known. Automatic learning then involves both network structure creation and conditional probability distributions estimation. Several algorithms of network learning are discussed in the literature, for example in Lucas et al. (2004).

4 Software

There are several options for a useful software to deal with graphical models. The most common packages are `Genie`, Hugin, BUGS and `R`. A very brief overview of the `Genie` software is presented in section 4.2, while the full manual can be found online at http://genie.sis.pitt.edu/wiki/GeNIe_Documentation. Several manuals for analysing Bayesian networks in `R` are also available, see e.g. Bøttcher et al. (2003a); Bøttcher et al. (2003b); or Scutari (2010).

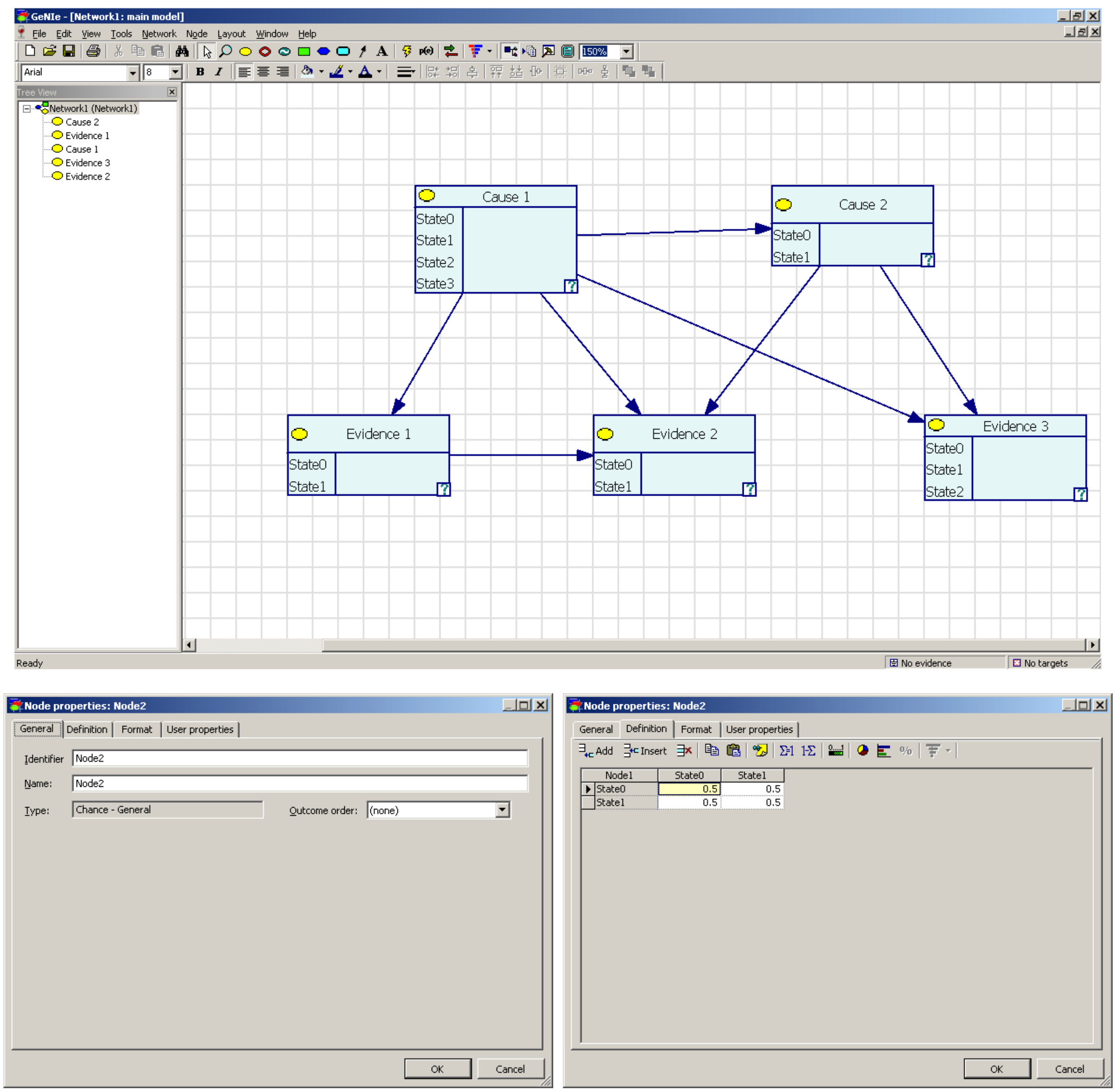
4.1 How to prepare data?

Graphical models use a conditional probability distribution at each node in the graph. If the conditional probability distribution is not known, it can be obtained from data by estimating the empirical conditional probability distribution (conditional frequencies). In case of automatic learning, all the relevant variables have to be organized in a single database structure. The software programs mentioned above can learn Bayesian networks from a `.dat`, `.txt`, `.csv`, or ODBC file. If the database is in a different format (e.g. Microsoft Access or SAS), the corresponding default software can usually translate the data-file into one of the readable formats.

4.2 Genie environment

The Genie software is a freeware and can be downloaded from http://genie.sis.pitt.edu. Genie has been designed for Windows platform PCs. However, it is possible to run Genie on a Mac using a program such as "Wine for Mac" emulator. The first step in manual designing a probabilistic network is to include all the nodes (random variables) by using the icon of a *yellow oval* from the tool-bar. The next step is to connect the nodes using the *arrow* icon from the toolbar to define probabilistic dependence between several pairs of nodes. A useful feature is to use the Node -> View as -> Bar Chart option. The result should look similar to figure 6.

Figure 6: Genie environment.



To specify characteristics of a node, double-click on the particular node. A window as in figure 6 should appear. In the `General` tab, name and identifier of the node can be defined. In the `Definition` tab, one can specify the conditional probability distribution at this node. Using the *thunder* icon, or the option `Network -> Update Immediately` reveals the prior probability distribution. Once evidence is obtained, clicking on the corresponding state of a node recalculates the posterior probabilities.

For automatic learning, the underlying database has to be imported into the program by `File -> Open Data File...` or `File -> Import ODBC data...`. The preferred algorithm may be selected under the option `Network -> Algorithm`. Additional features of the `Genie` package include for example a sensitivity analysis, showing strength of influence, or calculating probability of total evidence.

5 Suggested reading

Cowel et al. (1999) is an **essential read for beginners** to Bayesian networks. Especially chapters 2-4 provide a very clear, comprehensible theoretical introduction into the method illustrated with various examples. As an alternative, one may find the first chapter in Neapolitan (2003) also very helpful. Murphy (1998), Spiegelhalter (2004) and Airoldi (2007) present a brief overview of Bayesian networks; neither of these papers can be recommended as a source for deep understanding of the concept, but rather for getting some feeling what Bayesian networks are about.

Lucas et al. (2004) may be considered as a primary source for **practical construction** of Bayesian networks. The paper provides an overview of issues with both manual construction and automatic learning of Bayesian networks. Further discussion can be found in Neapolitan (2003). Few other advanced papers, e.g. Bøttcher et al. (2004) or Heckerman et al. (1994), focus on learning Bayesian networks, but these can be recommended only to experienced readers.

5.1 Studies using Bayesian networks

Bayesian networks may be applied in a wide range of areas in health services research (health economic evaluation, health quality measurement, health outcomes monitoring, cost-effectiveness analysis), but also in epidemiology, clinical research, medical decision making, or public health. The following list mentions several available studies which used Bayesian networks as their primary tool for modelling.

Decision making Medical decision making modelling; Acid (2004); Lucas (2001); Lucas et al. (2004).

Detecting errors An alternative method to detect blood lab errors better than the existing automated models; Doctor, Strylewicz (2010).

Economic evaluation Evaluation of quality-adjusted life years (QALYs) when health utilities are not directly available; cost-effectiveness analysis of an expensive and newly approved cancer drug; Quang (2010).

Epidemiology Modelling of the joint distribution of socio-demographic factors and obesity related behaviour; Harding (2011).

Mapping of measures Mapping health-profile or disease-specific measures onto preference-based measures; Quang (2010).

Outcomes monitoring Modelling of whether a new UK policy, which increased cervical cancer screening adherence, was associated with the observed decline in the incidence of cervical cancer; Spiegelhalter (1998).

Profit maximization Analysis and solving profit maximization problems from economics; Cobb (2011).

References

[1] Acid S., de Campos L. M., Fernández-Luna J. M., Rodríguez S., Rodríguez J. M., Salcedo J. L. (2004): *A Comparison of Learning Algorithm for Bayesian Networks: A Case Study Based on Data from an Emergency Medical Service.* Artificial Intelligence in Medicine 30, p. 215-232.

[2] Airoldi E. M. (2007): *Getting started in Probabilistic Graphical Models.* PLoS Computational Biology 3(12), 2421-2425.

[3] Bøttcher S. G., Dethlefsen C. (2003a): `deal`*: A Package for Learning Bayesian Networks.*

[4] Bøttcher S. G., Dethlefsen C. (2003b): *Learning Bayesian Networks with* `R`. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.

[5] Bøttcher S. G. (2004): *Learning Bayesian Networks with Mixed Variables.* Dissertation. Department of Mathematical Sciences, Aalborg University.

[6] Cobb B. (2011): *Graphical Models for Economic Profit Maximization.* Informs Transactions on Education 11(2), p. 43–56.

[7] Cowel R. G., Dawid A. P., Lauritzen S. L., Spiegelhalter D. J. (1999): *Probabilistic Networks and Expert Systems.* Springer-Verlag New York. ISBN 0-387-98767-3.

[8] Doctor J. N., Strylewicz G. (2010): *Detecting 'wrong blood in tube' errors: Evaluation of a Bayesian network approach.* Artificial Intelligence in Medicine 50, p. 75–82.

[9] Harding N. J. (2011): *Application of Bayesian Networks to Problems within Obesity Epidemiology.* Dissertation. Faculty of Medical and Human Sciences, University of Manchester.

[10] Heckerman D., Geiger D., Chickering D. M. (1994): *Learning Bayesian Networks: The Combination of Knowledge and Statistical Data.* Technical Report, Microsoft Research, MSR-TR-94-09.

[11] Jensen F. V. (1996): *An Introduction to Bayesian Networks.* Springer, ISBN 978-038-7915029.

[12] Lucas P. (2001): *Bayesian Networks in Medicine: a Model-based Approach to Medical Decision Making.*

[13] Lucas P. J. F., van der Gaag L. C., Abu-Hanna A. (2004): *Bayesian networks in biomedicine and health-care.* Artificial Intelligence in Medicine 30, p. 201–214.

[14] Madsen A. L., Jensen F., Kjaerulff U. B., Lang M. (2005): *The Hugin Tool for Probabilistic Graphical Models.* International Journal of Artificial Intelligence Tools 14(3), p. 507-543.

[15] Murphy K. (1998): *A Brief Introduction to Graphical Models and Bayesian Networks.*

[16] Neapolitan R. E. (2003): *Learning Bayesian Networks.* Prentice Hall, ISBN 978-013-0125347.

[17] Quang A. L. (2010): *New Approaches Using Probabilistic Graphical Models in Health Economics and Outcomes Research.* Dissertation. Faculty of the USC Graduate School, University of Southern California.

[18] Scutari M. (2010): *Learning Bayesian Networks with the* `bnlearn` `R` *Package.* Journal of Statistical Software.

[19] Spiegelhalter D. J. (1998): *Bayesian Graphical Modelling: A Case-Study in Monitoring Health Outcomes.* Applied Statistics 47, Part 1, p. 115-133.

[20] Spiegelhalter D. J., Abrams K. R., Myles J. P. (2004): *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Wiley. ISBN 978-0470-092590.

A Glossary

Arc See *edge*.

Bayesian network A model visually representing the joint probability distribution of a set of random variables by means of a *directed acyclic graph* and conditional probability distributions for each *node* in the graph.

Belief network See *Bayesian network*.

Branch See *edge*.

Directed acyclic graph A set of *nodes* and *directed edges*, which does not contain any cycle (i.e. it is not possible to get from one *node* back to itself, when following the *directed edges*).

Directed edge An *edge* with specified direction, which represents causal relationship between the two connected *nodes*.

Edge A representation of a conditional statistical dependence between a pair of *nodes* in a *Bayesian network*.

Graphical model See *Bayesian network*.

Learning a Bayesian network A method of automatic construction of a Bayesian network from a database using an appropriate software.

Likelihood function A retrospective probability of the observed data.

Node A representation of a random variable in a *Bayesian network*.

Odds A ratio of the probability that an event will happen to the probability that it will not happen.

Posterior distribution A probability distribution of a random variable composed of the *prior distribution* and the *likelihood function* of the data.

Prior distribution A probability distribution assigned to a random variable before the incorporation of data.

Probabilistic network See *Bayesian network*.