Bayesian Networks

Daniel Otero Fadul

Science Department
School of Engineering and Science

Bayesian Networks

A **Bayesian network** is a structure that allows us to represent dependencies between different random variables. These types of networks can represent virtually any joint distribution, often in a very concise manner. This representation is done using a **directed acyclic graph**.

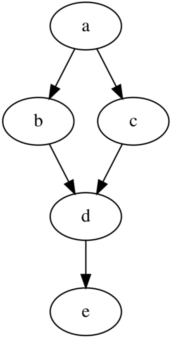

Figura: A **directed acyclic graph** is a type of graph in which its edges have a direction and which does not have cycles that establish a closed path between a node, or vertex, and itself. Image taken from taken from `https://en.wikipedia.org/wiki/Directed_acyclic_graph`).

Bayesian Networks

In the context of Bayesian networks, these can be considered directed acyclic graphs that meet the following characteristics:

- Each node is a random variable, which can be discrete or continuous.
- A set of directed arcs connect pairs of nodes. If there is an arc going from node x to node y, x is said to be the *parent* of y.
- Each node x_i that has *parents* has an associated conditional probability distribution $P(x_i|\text{parents}(x_i))$.

Bayesian Networks

The network topology, the set of nodes and arcs, specifies the way in which cause-and-effect relationships are established between variables: causes are the parents of effects. These relationships are normally established by an expert.

Once the topology of the Bayesian network has been established, it is possible to determine the conditional probability distributions of each variable given its parents. The combination of the probability distributions of each node is sufficient to obtain the joint probability distribution of all variables.

As we know, thanks to the multiplication rule, any joint probability distribution can be written as follows:

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= P(x_1)P(x_2|x_1)\cdots P(x_n|x_{n-1}, x_{n-2}, \ldots, x_1) \\ &= P(x_1)\prod_{i=2}^{n} P(x_i|x_{i-1}, \ldots, x_1). \end{aligned}$$

In the context of Bayesian networks, this same joint probability distribution would be equal to

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(x_i)).$$

This is valid as long as $\text{parents}(x_i) \subseteq \{x_{i-1}, x_{i-2}, \ldots, x_1\}$. This condition can be satisfied if the enumeration of the nodes is consistent with the implicit order of the graph structure. Incidentally, if the random variable x_i has no parents, we have that $P(x_i|\text{parents}(x_i)) = P(x_i)$.

Expressing the joint probability distribution as the multiplication of the conditional probabilities mentioned above is possible if the following condition is met:

$$P(x_i|x_{i-1},\ldots,x_1) = P(x_i|\text{parents}(x_i)). \tag{1}$$

This condition allows the Bayesian network to be a correct representation of the joint probability distribution if each node is conditionally independent of its predecessors given its parents. The following methodology guarantees this condition:

1. **Nodes:** Determine all the variables in the model. Once this is done, list them in such a way that the causes precede the effects.
2. **Arcs:** For each x_i:
   - From the subset $\{x_{i-1}, x_{i-2}, \ldots, x_1\}$, choose a minimal set of parents such that Equation (1) is satisfied.
   - Connect each parent by means of a link with x_i.
   - Determine the **tconditional distribution table** $P(x_i|\text{parents}(x_i))$.

An Example

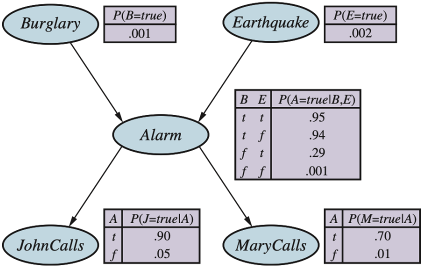

Figura: An example of a Bayesian network showing its topology and conditional distribution tables. In the tables, the letters B, E, A, J, and M refer to "Burglary," "Earthquake," "Alarm," "JohnCalls," and "MaryCalls," respectively. Image taken from [1].

Note that $P(\text{MaryCalls}|\text{Burglary, Alarm}) = P(\text{MaryCalls}|\text{Alarm})$. On the other hand, the probabilities of each row of the tables must add up to one, since all the variables are binary, the values that would be taken by the conditional distributions for the complements of events B, E, A, J, and M are omitted. Once the network has been determined, we can do things like the following:

$$\begin{aligned} P(\bar{b}, \bar{e}, a, j, m) &= P(\bar{b})P(\bar{e})P(a|\bar{b}, \bar{e})P(j|a)P(m|a) \\ &= (0.999)(0.998)(0.001)(0.9)(0.7) = 0.000628. \end{aligned}$$

Monty Hall Problem

Figura: A scene in which Amy Santiago discusses the "Monty Hall" problem in the series "Brooklyn 99." Image taken from
`https://thelachatupdate.com/2018/11/06/the-monty-hall-problem-simplified/`.

Monty Hall Problem

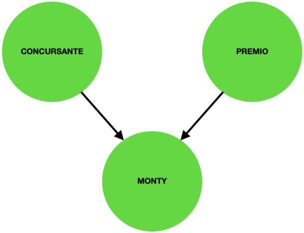

Figura: The Bayesian network of the "Monty Hall" problem.

Given what we have seen, the random variables in this problem are "Contestant," "Prize," and "Monty," which we denote as C, P, and M, respectively. Let 1, 2, and 3 be the doors in the "Monty Hall" problem; these numbers are the values that the random variables take. Note that the random variables C and P are independent and that the variable M depends on C and P.

Monty Hall Problem

We have that

$$\begin{aligned} P(C,P,M) &= P(C)P(P|C)P(M|C,P) \\ &= P(C)P(P)P(M|C,P). \end{aligned}$$

Also, the tables of $P(C)$ y $P(P)$ are as follows:

	$C=1$	$C=2$	$C=3$
$P(C)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

	$P=1$	$P=2$	$P=3$
$P(P)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Monty Hall Problem

As for the table of $P(M|C,P)$, we have that

	$M=1$	$M=2$	$M=3$
$P(M\|C=1,P=1)$	0	$\frac{1}{2}$	$\frac{1}{2}$
$P(M\|C=1,P=2)$	0	0	1
$P(M\|C=1,P=3)$	0	1	0
$P(M\|C=2,P=1)$	0	0	1
$P(M\|C=2,P=2)$	$\frac{1}{2}$	0	$\frac{1}{2}$
$P(M\|C=2,P=3)$	1	0	0
$P(M\|C=3,P=1)$	0	1	0
$P(M\|C=3,P=2)$	1	0	0
$P(M\|C=3,P=3)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Monty Hall Problem

To solve the "Monty Hall" problem, it is useful to obtain the distribution of $P(P|C=1, M=2)$:

$$\begin{aligned}
P(P|C=1, M=2) &= \frac{P(C=1)P(P)P(M=2|C=1,P)}{P(C=1, M=2)} \\
&= \frac{P(C=1)P(P)P(M=2|C=1,P)}{P(C=1)P(M=2|C=1)} \\
&= \frac{\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)P(M=2|C=1,P)}{\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)} \\
&= \frac{2}{3}P(M=2|C=1,P).
\end{aligned}$$

Therefore,

	$P=1$	$P=2$	$P=3$
$P(P\|C=1, M=2)$	$\frac{1}{3}$	0	$\frac{2}{3}$

Example of an Exact Inference

Returning to the example of Mary and John, suppose we now want to calculate $P(B|J=t, M=t)$. If this is the case, we would have to do the following

$$\begin{aligned} P(B|J=t, M=t) &= \alpha P(B, J=t, M=t) \\ &= \alpha \sum_E \sum_A P(B, J=t, M=t, E, A) \\ &= \alpha \sum_E \sum_A P(B)P(E)P(A|B,E)P(J=t|A)P(M=t|A), \end{aligned}$$

where α is a constant. It can be seen that, although the Bayesian network is not very complex, many calculations are required to obtain the conditional probability distribution $P(B|J=t, M=t)$. In fact, for n Boolean variables, the complexity of this procedure is $O(n2^n)$.

Computational Limitations

The latter is an example of **exact inference** for a Bayesian network. It is obviously accurate, but computationally very expensive. In general, the algorithms used to perform these calculations can be divided into two categories: those that perform an exact inference and those that obtain an **approximate inference**. For more information, please refer to [1].

BIBLIOGRAFÍA

1 Stuart Rusell,Peter Norvig, S. J., *"Artificial Intelligence: A Modern Approach"*, Cuarta Edición, Prentice Hall, 2020.

2 Charniak, E. *"Bayesian Networks without Tears"*, AI Magazine, 12(4), 50, 1991.