Casual and Bayesian networks

Chapter 2

Examples

2.1 Key roads

In this section, we give two examples. They illustrate crucial points to consider when reasoning about certainty in the field.

2.1.1 Key roads

This chapter introduces causal networks as graphical representations of casual relations among variables. Bayesian networks and the reasoning is performed through probability modelled as Bayesian networks and the reasoning is performed through probability modelled as Bayesian networks and the reasoning is performed through probability modelled as Bayesian networks and the reasoning is performed through probability modelled as Bayesian networks and the reasoning is performed through probability.
2.1.4 Parthenogenesis of the Galaxy

In your network of belief, the world's events, perfectly timed and closely connected, are a foundation for understanding. This is from a foundational point of view, perfectly timed, perfectly connected, perfectly understood. The models of the world have no more mystery between them. The models are connected by a chain of reasoning that is used for the 'reasoning' aspect of calculation. The reason is used for the reasoning aspect of calculation. The reason is used for the reasoning aspect of calculation.

The event is the result of the reasoning drawn on the model. This is from a foundational point of view, perfectly timed, perfectly connected, perfectly understood.

2.1.3 Causation and Reasoning

Figure 2.2: A network model of wet grass.}

H

S

R

W

Wet grass is a cause of moisture, but it also causes the grass to be wet. This shows that the grass is wet because it is raining.

Hypsématon is called conditional independence. The event is the effect of the reasoning drawn on the model. This is from a foundational point of view, perfectly timed, perfectly connected, perfectly understood. The event is the result of the reasoning drawn on the model. This is from a foundational point of view, perfectly timed, perfectly connected, perfectly understood.
From the remaining un-instrumented variables: (a) if $A$ is d-separated from $B$, $D$, $E$ and $F$ in $G$ and if $A$, $B$, $D$, $E$ and $F$ are all unconnected, then $G$, $H$, $I$, $J$, $K$, $L$, $M$, and $N$ are all unconnected (b) although $G$ may be connected, if $A$, $B$, $D$, $E$, and $F$ are all unconnected, then $G$, $H$, $I$, $J$, $K$, $L$, $M$, and $N$ are all unconnected.

**Figure 2.9** Causal networks with hard evidence cutset (the $\text{d}$-separated).

**Figure 2.2**

- **Definition (d-separation):** Two variables $A$ and $B$ in a causal network are d-separated through a variable and conditioning the value of any parent of $A$ or $B$, then $A$ and $B$ are d-separated.

- **2.2.1 d-separation**

The d-separation of evidence

**Figure 2.7** Examples where the parents of $A$ are dependent.

- **Remark:** Evidence on a variable is a remnant of the causalities of his parents. If $A$ and $B$ are not d-separated, we call them d-connected.

- **Figure 2.8** A causal network with $A$ and $B$ unconnected.

- **Figure 2.6** Causal network.

- **Figure 2.5** Causal network.

- **Figure 2.4** Causal network.

- **Figure 2.3** Causal network.

- **Figure 2.2** Causal network.

- **Figure 2.1** Causal network.

- **Figure 2.0** Causal network.
A conditional probability statement is the following:

\[ P(A | B) = \frac{P(A \cap B)}{P(B)} \]

Sometimes \( P(A | B) \) is called the likelihood of \( A \) given \( B \), and is denoted \( L(A | B) \).

In the Bayesian formalism or causal assessment, the probability of \( A \) given \( B \) is derived from the likelihood of \( A \) given \( B \), the prior probability of \( B \), and the evidence value of \( B \).
(1) if more than 2,000 copies are sold in 1997 I will receive $100;

(2) I will by the end of 1997 be allowed to draw a ball from an urn with \( n \) balls and 100 \(- n\) white balls. If my ball is red I will get $100.

Now, if all balls in the urn are red I will prefer (2), and if all balls are white I will prefer (1). There is a number \( n \) for which the two gambles are equally attractive, and for this \( n, \frac{n}{100} \) is my estimate of the probability of selling more than 2,000 copies of this book in 1997 (I shall not disclose the \( n \) to the reader).

For subjective probabilities defined through such ball drawing gambles the fundamental rule can also be proved.

2.3.4 Probability calculus for variables

As stated in Section 2.2, the nodes in a causal network are variables with a finite number of mutually exclusive states.

If \( A \) is a variable with states \( a_1, \ldots, a_n \), then \( P(A) \) is a probability distribution over these states:

\[
P(A) = (x_1, \ldots, x_n) \quad x_i \geq 0 \quad \sum_{i=1}^{n} x_i = 1,
\]

where \( x_i \) is the probability of \( A \) being in state \( a_i \).

**Notation.** The probability of \( A \) being in state \( a_i \) is denoted \( P(A = a_i) \) and denoted \( P(a_i) \) if the variable is obvious from the context.

If the variable \( B \) has states \( b_1, \ldots, b_m \), then \( P(A \mid B) \) is an \( n \times m \) table containing numbers \( P(a_i \mid b_j) \) (see Table 2.1).

\( P(A, B) \), the joint probability for the variables \( A \) and \( B \), is also an \( n \times m \) table. It consists of a probability for each configuration \( (a_i, b_j) \) (see Table 2.2).

When the fundamental rule (2.1) is used on variables \( A \) and \( B \), then the procedure is to apply the rule to the \( n \cdot m \) configurations \( (a_i, b_j) \):

\[
P(a_i \mid b_j)P(b_j) = P(a_i, b_j).
\]

This means that in the table \( P(A \mid B) \), for each \( j \) the column for \( b_j \) is multiplied by \( P(b_j) \) to obtain the table \( P(A, B) \). If \( P(B) = (0.4, 0.4, 0.2) \) then Table 2.2 is the result of using the fundamental rule on Table 2.1. When applied to variables, we use the same notation for the fundamental rule:

\[
P(A \mid B)P(B) = P(A, B).
\]

From a table \( P(A, B) \) the probability distribution \( P(A) \) can be calculated. Let \( a_i \) be a state of \( A \). There are exactly \( m \) different events for which \( A \) is in state \( a_i \), namely the mutually exclusive events \( (a_i, b_1), \ldots, (a_i, b_m) \). Therefore, by axiom (ii)

\[
P(a_i) = \sum_{j=1}^{m} P(a_i, b_j).
\]

### Table 2.1 An example of \( P(A \mid B) \).

<table>
<thead>
<tr>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>0.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Note that the columns sum to one.

### Table 2.2 An example of \( P(A, B) \).

<table>
<thead>
<tr>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>0.24</td>
<td>0.28</td>
</tr>
</tbody>
</table>

This calculation is called marginalization and we say that the variable \( B \) is marginalized out of \( P(A, B) \) (resulting in \( P(A) \)). The notation is

\[
P(A) = \sum_{B} P(A, B).
\]

By marginalizing \( B \) out of Table 2.2 we get \( P(A) = (0.4, 0.6) \).

The division in Bayes' rule (2.3) is treated in the same way as the multiplication in the fundamental rule (see Table 2.3).

2.3.5 Conditional independence

The blocking of transmission of evidence as described in Section 2.2.1 is, in the Bayesian calculus, reflected in the concept of conditional independence. The variables \( A \) and \( C \) are independent given the variable \( B \) if

\[
P(A \mid B) = P(A \mid B, C).
\]

This means that if the state of \( B \) is known then no knowledge of \( C \) will alter the probability of \( A \).

### Table 2.3 \( P(B \mid A) \) as a result of applying Bayes' rule to Table 2.1 and \( P(B) = (0.4, 0.4, 0.2) \).

<table>
<thead>
<tr>
<th>( b_1 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_2 )</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>( b_3 )</td>
<td>0.3</td>
<td>0.47</td>
</tr>
<tr>
<td>( b_3 )</td>
<td>0.3</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Bayesian Networks

The chain rule

\[ p(A, B, C, D) = p(A) p(B|A) p(C|B) p(D|C) \]

Figure 2.11 A directed acyclic graph (DAG). The probabilities

2.2.6 Definition of Bayesian networks

So, why bother with the transmission of information? Suppose we receive a message that the radio is not working. However, our conditional independence assumption holds. Then, by the definition of independence, the radio is independent of the transmission.

Figure 2.10 Conditional independence appears in the case of causal and directed connections.
The information that a proposition is true is now used to update the probability of
\[ p(A) = (H) \cdot d = (M) \cdot d \]
and so on.

In order to get the probabilities for \( M \) and \( H \), we manipulate our data from Table 2.5

\[ \begin{array}{llll}
0.9 & 0.4 & 6.0 & 0.95 - 0.05 = 0.90 \times (I) \cdot d & (\text{for } \mathcal{P}(A) \cdot d)
\end{array} \]

Table 2.5 gives all 16 probabilities.

The fundamental rule is.

\[ \mathcal{P}(A) \cdot d = \mathcal{P}(A | \mathcal{P}(M)) \cdot d = \mathcal{P}(A | \mathcal{P}(H)) \cdot d = \mathcal{P}(A | \mathcal{P}(M) \wedge \mathcal{P}(H)) \cdot d \]

By the fundamental rule we have.

\[ \mathcal{P}(A) \cdot d = \mathcal{P}(A | \mathcal{P}(M)) \cdot d \]

From the introduction hypothesis we have that the product of all

\[ \text{Figure 2.13 A DAG with } n \text{ variables. If the variable } \mathcal{P}(M) \text{ is removed, the induction hypothesis can be applied.} \]

2.4.1 Key Roads

Calculations considerably easier than those in this section. Form the marginal probability of \( M \) and \( H \) by

\[ \begin{array}{llll}
0.9 & 0.4 & u & \mathcal{P}(M) \cdot d
\end{array} \]

\[ \begin{array}{llll}
0.9 & 0.05 & \mathcal{P}(H) \cdot d
\end{array} \]

\[ (\mathcal{P}(M) \cdot d, \mathcal{P}(H) \cdot d) \]

Table 2.5 Joint probabilities for \( M \) and \( H \).

Table 2.4 Conditional probabilities for \( H \) and \( M \).

\[ \begin{array}{llll}
6.0 & 2.0 & \mathcal{P}(M) \cdot d & \mathcal{P}(H) \cdot d
\end{array} \]

\[ \begin{array}{llll}
1.0 & 8.0 & \mathcal{P}(M) \cdot d & \mathcal{P}(H) \cdot d
\end{array} \]

\[ \mathcal{P}(M) \cdot d = \mathcal{P}(M) \cdot d \]

The examples revisited.
The distribution of \( (Y, | H) \) is calculated through marginalization of 
the joint probability mass function \((Y, W, H)\) for the sum of the elements in Table 2.1. 
By the sum of the elements in Table 2.1, we simply normalize the table by dividing 
the row sum of each row by the corresponding parameter \( (\lambda = H) \) and then divide by another 
representation of \( \lambda = H \) to get the probability mass function of \( \lambda = H \). Since the result shall be a probability mass 
function, it is used to update the previous mass function \( (Y, | H) \) and then mass function \( (Y, W, H) \).

**Figure 2.14** The clusters for the west east example. They

![Figure 2.14](image)

<table>
<thead>
<tr>
<th>( S )</th>
<th>( Y )</th>
<th>( W )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>H</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>H</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>H</td>
</tr>
</tbody>
</table>

\( u = S \), \( \lambda = W \), \( \kappa = H \)

**Table 2.7** The probabilities for the west east example. The

<table>
<thead>
<tr>
<th>( S )</th>
<th>( Y )</th>
<th>( H )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\( u = S \), \( \lambda = W \), \( \kappa = H \)

Finally, calculate the distribution of \((Y, | H)\) by marginalizing out on \( (Y, W, H) \).

**Table 2.6** Tables showing the calculation of \((Y, | H)\) as shown in Table 2.6.

To compute the probability of \( H \), first we use the fundamental rule (2.1) to calculate 

\( (Y, | H) d (\lambda = H) = \frac{(Y | \lambda = H) d (\lambda = H)}{(\lambda = H) d} = \frac{(Y, | H) d (\lambda = H)}{(\lambda = H) d} \)

For this page, the result is:

\( (1 | H) d = \frac{(Y, | H) d (\lambda = H)}{(\lambda = H) d} \)

The examples revised
By marginalizing we get $p(H) = 0.160$(see Table 2.13.2).

$\begin{align*}
\frac{(y)_{sd}}{(y)_{sd}d}\frac{(s', H, K, \theta')_{sd}d}{(s, H, K, \theta')_{sd}d} = 1
\end{align*}$

The calculation follows the same pattern. A message on $p(y)$ is sent from $p(y, m)$. The result must reflect we still have to calculate $p(y, m | S)$ (see Table 2.12). We get

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0.736 & 0 & 0.0528 & 0 & 0.649 \\
\hline
0 & 0.967 & 0 & 0.032 & 0 & 0.969 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
(0, 0) & (0, 0.669) & (0, 0.182) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) \\
\hline
(0, 0) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) & (0.96, 0.04) & (0.96, 0.04) \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
(0, 0) & (0, 0.669) & (0, 0.182) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) \\
\hline
(0, 0) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) & (0.96, 0.04) & (0.96, 0.04) \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
(0, 0) & (0, 0.669) & (0, 0.182) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) \\
\hline
(0, 0) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) & (0.96, 0.04) & (0.96, 0.04) \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
(0, 0) & (0, 0.669) & (0, 0.182) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) \\
\hline
(0, 0) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) & (0.96, 0.04) & (0.96, 0.04) \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
\end{align*}

\begin{align*}
\begin{array}{c|c|c|c|c|c}
(0, 0) & (0, 0.669) & (0, 0.182) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) \\
\hline
(0, 0) & (0.18, 0.02) & (0.18, 0.02) & (0.96, 0.04) & (0.96, 0.04) & (0.96, 0.04) \\
\hline
\end{array}
\end{align*}
The fundamental rule for probabilistic calculus:

\[ p(A | B, C, D) = p(A | C) p(D | B, C) \]

**Summary**

2.6 Summary

(largely containing the importance of evidence)

Because probabilistic models have not been discussed,

study of evidence from factors is essential to probabilistic models.

Evidence is entered into the variables factor and probabilistic.

![Diagram of probabilistic model](image)

**Figure 2.17**

The part of BOLTO modeling partial evidence.

Next, the impact of evidence from factors is considered in the node.

Because evidence has

varieties probabilistic models are known.

The network model in BOLTO also has part model.

![Diagram of network model](image)

**Figure 2.16**

Hypothesis of evidence from each parent one.

AND BAYESIAN NETWORK

The network model in BOLTO also has part model.

![Diagram of network model](image)

**Figure 2.15**

Hypothesis of evidence from each parent one.

CAUSAL AND BAYESIAN NETWORK

The network model in BOLTO also has part model.

![Diagram of network model](image)
Exercises 2.4 Let $D_1$ and $D_2$ be DAGs over the same variables. $D_1$ is an $I$-equivalent of $D_2$ if all d-separation properties of $D_1$ also hold for $D_2$. If $D_1$ is an I-equivalent of $D_2$, then $D_2$ is also an I-equivalent of $D_1$.

Show that $A$ is d-separated from the remaining unmeasured variables, where child of $A$ is denoted as $(A)$. The parents of $A$ are denoted as $\text{pa}(A)$.

Figure for Exercise 2.2

Exercises 2.2 Let $G$ be a variable in a DAG. Assume that the following variables are d-connected to $G$.

Admissibility of d-separation in Bayesian networks

To each variable $A$, assign the set of its parents, $\text{pa}(A)$. Then $A$ is d-separated from $B$. The variables together form a directed acyclic graph (DAG) and each variable has at least one set of parents.

A set of variables that form a directed edges between variables.

A Bayesian network consists of the following:

Definition of Bayesian networks

A and $C$ are independent given $B$. The independence of $B$.

Conditional Independence

$$p(A, B, C) = p(A) p(B | A) p(C | B) = p(A) p(C | B) = p(A) p(B) p(C | B)$$

Marginalization

$$p(C | A) p(B | C, A) = p(B | A)$$

Bayes' Rule

$$p(A | B) = \frac{p(B | A) p(A)}{p(B)}$$

Bibliographic notes

27 Causal and Bayesian networks
Exercises 2.8
Perform a Bayesian calculation of the reasoning in Section 2.1.4 (causal network).

Calculate $p(A, B, C)$.

Table 2.16 Conditional probability tables for Exercise 2.6.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$p(A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.21</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 2.17 Tables for Exercise 2.8. Probabilities for radio alarm.

Table 2.15 Conditional probability tables for Exercise 2.4.

Exercise 2.9 Let $p(C | A, B)$.

Given $B$ and only if $p(A, C | B) = p(B) = p(A) = 0.001$. Prove that $A$ and $C$ are independent.

Exercises 2.7
Calculate $p(A, B, C)$ given in Table 2.16.

Figure for Exercise 2.4.

Exercise 2.5 Calculate $p(A, C | p(B)$, $p(A | B)$, and $p(A)$ from Table 2.14.

Table 2.14 Tables for Exercise 2.5.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$C$</th>
<th>$p(A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.005</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.005</td>
</tr>
</tbody>
</table>