Practical Lecture 10 - Bayesian Networks

Andreas Wichert and Luis Sa-Couto
INESC-ID & Instituto Superior Técnico, Universidade de Lisboa
{luis.sa.couto,andreas.wichert}@tecnico.ulisboa.pt

1 Bayesian Network

For binary events there are two states: true or false. In other words, for any event a there is an event $\neg a$ which corresponds to the event that a does not occur. Binary events are described by binary variables. For binary events,

$$p(x) + p(\neg x) = 1, \quad p(y) + p(\neg y) = 1, \tag{1}$$

where the law of total probability is represented by

$$p(y) = p(y, x) + p(y, \neg x) = p(y|x) \cdot p(x) + p(y|\neg x) \cdot p(\neg x) \tag{2}$$

and

$$p(\neg y) = p(\neg y, x) + p(\neg y, \neg x) = p(\neg y|x) \cdot p(x) + p(\neg y|\neg x) \cdot p(\neg x). \tag{3}$$

The relationship can be represented by a graph (see Figure 1) that indicates the influence between events x and y. If two events x and y are independent, then the probability that events x and y both occur is

$$p(x, y) = p(x \wedge y) = p(x) \cdot p(y). \tag{4}$$

In this case the conditional probability is

$$p(x|y) = p(x). \tag{5}$$

If all n possible variables are independent, then

$$p(x_1, x_2, \cdots, x_n) = p(x_1) \cdot p(x_2) \cdot \cdots \cdot p(x_n) = \prod_{i=1}^{n} p(x_i) \tag{6}$$

In the case that not all variables are independent we can decompose the probabilistic domain into subsets via conditional independence. For a subset of dependent variables

$$p(x_1, x_2) = p(x_1|x_2) \cdot p(x_2) = p(x_2|x_1) \cdot p(x_1). \tag{7}$$

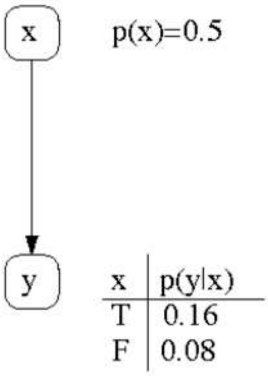


Figure 1: The causal relation between events x and y represented by a direct graph of two nodes. Note that each node is followed by a conditional probability table that specifies the probability distribution of that node according to its parent node. This direct graph with two nodes representation corresponds to a simple Bayesian network.

This follows from the Bayes' rule

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{p(x_2|x_1) \cdot p(x_1)}{p(x_2)} \tag{8}$$

Two variables x_1 and x_2 are conditionally independent given x_3 if

$$p(x_1|x_2, x_3) = p(x_1|x_3). \tag{9}$$

Assuming x_2 and x_3 are independent, but x_1 is conditionally dependent given x_2 and x_3 then

$$p(x_1, x_2, x_3) = p(x_1|x_2, x_3) \cdot p(x_2) \cdot p(x_3). \tag{10}$$

Assuming x_4 is conditionally dependent given x_1 but independent of x_2 and x_3 then

$$p(x_1, x_2, x_3, x_4) = p(x_1|x_2, x_3) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4|x_1). \tag{11}$$

This relationship between occurrence of events called causality is represented by conditional dependency.. The relationship can be represented by a graph (see Figure 2) that indicates the influence between events x_1, x_2, x_3 and x_4. In our example x_2 and x_3 cause x_1 and only then x_1 causes x_4. This kind of decomposition via conditional independence is modelled by Bayesian networks. Bayesian networks provide a natural representation for (causally induced) conditional independence. They represent a set of conditional independence assumptions by the topology of an acyclic directed graph and sets of conditional probabilities. In the network each variable is represented by a node and the links between them represent the conditional independence of the variable towards its non descendants and its immediate predecessors (see Figure 2). The corresponding network topology reflects our belief in the associated causal knowledge. Consider the well-known example of Judea Perl "I am at work in Los Angeles, and neighbour John calls to say that the alarm of my house is ringing. Sometimes minor earthquakes set off the alarm. Is there a burglary?" Constructing a Bayesian network , should be easy because each variable is directly influenced by only a few other variables. In the example, there are four variables, namely, $Burglary(= x_2)$, $Earthquake(= x_3)$, $Alarm(= x_1)$ and $JohnCalls(= x_4)$. Due to simplicity, we ignore an additional variable $MaryCalls$ that was present in the original example. The corresponding network topology in Figure 2 reflects the following "causal" knowledge:

- A burglar can set the alarm on.
- An earthquake can set the alarm on.
- The alarm can cause John to call.

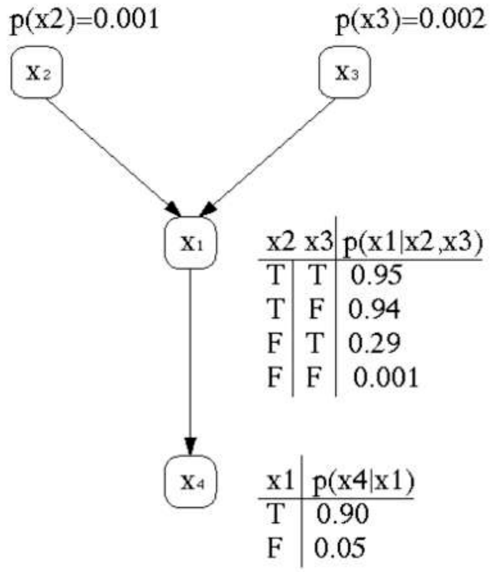


Figure 2: A Bayesian network representing the causal relationship between events x_1, x_2, x_3 and x_4. The four variables can be associated with causal knowledge, in our example $Burglary(= x_2)$, $Earthquake(= x_3)$, $Alarm(= x_1)$ and $JohnCalls(= x_4)$.

Bayesian networks represent for each variable a conditional probability table which describes the probability distribution of a specific variable given the values of its immediate predecessors. A conditional distribution for each node x_i given its parents is

$$p(x_i|Parent_1(x_i), Parent_2(x_i), .., Parent_i(x_i))$$

with k representing the number of predecessor nodes (or parent nodes) of node x_i

$$p(x_1, .., x_n) = \prod_{i=1}^{n} p(x_i|Parent_1(x_i), Parent_2(x_i), .., Parent_i(x_i)). \tag{12}$$

Given the query variable x whose value has to be determined and the evidence variable e which is known and the remaining unobservable variables y, we perform a summation over all possible y. In the following examples, for simplification the variables are binary and describe binary events. All possible values (true/false) of the unobservable variables y are determined according to the law of total probability

$$p(x|e) = \alpha \sum_{y} p(x, e, y) = \alpha \cdot (p(x, e, y) + p(x, e, \neg y)). \tag{13}$$

or

$$p(x|e) = \alpha \sum_{y} p(x, e, y) = \alpha \cdot (p(x, e|y) \cdot p(y) + p(x, e|\neg y) \cdot p(\neg y)). \tag{14}$$

with

$$\alpha = \frac{1}{p(e)} = \frac{1}{\sum_y p(x, e, y) + \sum_y p(\neg x, e, y)} \tag{15}$$

and

$$1 = \alpha \cdot \left(\sum_{y} p(x, e, y) + \sum_{y} p(\neg x, e, y) \right).$$

For the preceding example

$$p(x_4|x_1, x_2, x_3) = \alpha \cdot p(x_1|x_2, x_3) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4|x_1) \tag{16}$$

for an unknown variable, for example x_3, we will use the notion X_3 to indicate that its value is unknown. We can then apply the law of total probability by making a summation over its values.

$$p(x_1, x_2, X_3, x_4) = p(x_1, x_2, x_3, x_4) + p(x_1, x_2, \neg x_3, x_4) \tag{17}$$

$$p(x_1, x_2, X_3, x_4) = p(x_2) \cdot p(x_4|x_1) \cdot \left(\sum_{x_3} p(x_1|x_2, x_3) \cdot p(x_3) \right). \tag{18}$$

$$p(x_4|x_1, x_2) = \alpha \cdot p(x_2) \cdot p(x_4|x_1) \cdot \left(\sum_{x_3} p(x_1|x_2, x_3) \cdot p(x_3) \right). \tag{19}$$

$$p(x_4|x_1,x_2) = \alpha \cdot p(x_2) \cdot p(x_4|x_1) \cdot (p(x_1|x_2,x_3) \cdot p(x_3)+$$
$$+p(x_1|x_2,\neg x_3) \cdot p(\neg x_3)) \tag{20}$$

with

$$\alpha = \frac{1}{p(x_1,x_2)} = \frac{1}{p(x_1,x_2,X_3,x_4)+p(x_1,x_2,X_3,\neg x_4)} \tag{21}$$

and

$$p(x_1,x_2) = p(x_1,x_2,X_3,x_4)+p(x_1,x_2,X_3,\neg x_4). \tag{22}$$

After calculating, we arrive at

$$p(x_4|x_1,x_2) = \frac{p(x_4|x_1,x_2)}{p(x_4|x_1,x_2)+p(\neg x_4|x_1,x_2)}. \tag{23}$$

For unknown variables x_1, x_3 indicated by X_1 andX_3, we apply the law of total probability.

$$p(X_1,x_2,X_3,x_4) = p(x_1,x_2,x_3,x_4)+p(\neg x_1,x_2,x_3,x_4)+$$
$$+p(x_1,x_2,\neg x_3,x_4)+p(\neg x_1,x_2,\neg x_3,x_4) \tag{24}$$

$$p(X_1,x_2,X_3,x_4) = p(x_2) \cdot \sum_{x_1} \left(p(x_4|x_1) \cdot \left(\sum_{x_3} p(x_1|x_2,x_3) \cdot p(x_3) \right) \right). \tag{25}$$

$$p(x_4|x_2) = \alpha \cdot p(x_2) \cdot (p(x_4|x_1) \cdot p(x_1|x_2,x_3) \cdot p(x_3)+$$
$$+p(x_4|x_1) \cdot p(x_1|x_2,\neg x_3) \cdot p(\neg x_3)+$$
$$+p(x_4|\neg x_1) \cdot p(\neg x_1|x_2,x_3) \cdot p(x_3)+p(x_4|\neg x_1) \cdot p(\neg x_1|x_2,\neg x_3) \cdot p(\neg x_3)) \tag{26}$$

with

$$\alpha = \frac{1}{p(x_2)} = \frac{1}{p(X_1,x_2,X_3,x_4)+p(X_1,x_2,X_3,\neg x_4)}. \tag{27}$$

1.1 Example

As indicated in Figure 2 determine the probability of x_1 given x_4 and x_5 are false and x_2 is unknown.

$$p(x_1|\neg x_4,\neg x_3) = \alpha \cdot p(x_1,\neg x_4,\neg x_3)$$

$$p(x_1,\neg x_4,\neg x_3) = \sum_{x_2} p(x_2) \cdot p(\neg x_3) \cdot p(x_1|x_2,\neg x_3) p(\neg x_4|x_1)$$

$$p(x_1,\neg x_4,\neg x_3) = p(\neg x_3) \cdot p(\neg x_4|x_1) \cdot \left(\sum_{x_2} p(x_2) \cdot p(x_1|x_2,\neg x_3) \right)$$

$$p(x_1,\neg x_4,\neg x_3) = p(\neg x_3) \cdot p(\neg x_4|x_1) \cdot (p(x_2) \cdot p(x_1|x_2,\neg x_3)+$$
$$+p(\neg x_2) \cdot p(x_1|\neg x_2,\neg x_3))$$

$$p(x_1, \neg x_4, \neg x_3) = (1 - 0.002) \cdot (1 - 0.9) \cdot (0.001 \cdot 0.94 + (1 - 0.001) \cdot 0.001)$$

$$p(x_1, \neg x_4, \neg x_3) = 0.000193512.$$

$$p(\neg x_1, \neg x_4, \neg x_3) = p(\neg x_3) \cdot p(\neg x_4 | \neg x_1) \cdot (p(x_2) \cdot p(\neg x_1 | x_2, \neg x_3) +$$

$$+ p(\neg x_2) \cdot p(\neg x_1 | \neg x_2, \neg x_3))$$

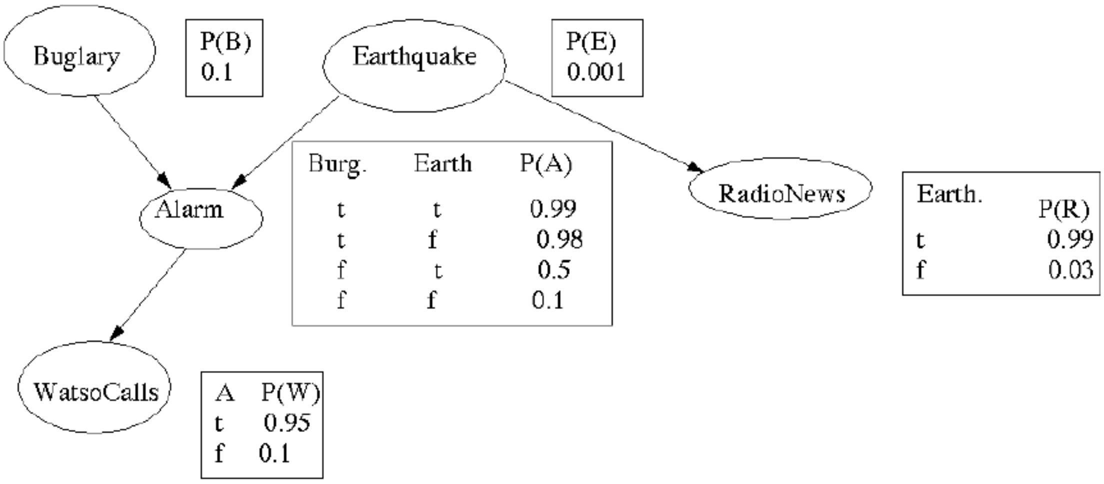
$$p(\neg x_1, \neg x_4, \neg x_3) = (1 - 0.002) \cdot (1 - 0.05) \cdot$$

$$\cdot (0.00.1 \cdot ((1 - 0.94) + 1 - 0.001) \cdot (1 - 0.001))$$

$$p(\neg x_1, \neg x_4, \neg x_3) = 0.946205.$$

$$p(x_1 | \neg x_4, \neg x_3) = \frac{p(x_1, \neg x_4, \neg x_3)}{p(x_1, \neg x_4, \neg x_3) + p(\neg x_1, \neg x_4, \neg x_3)} = 0.000204472.$$

1) Belief Networks



(a) Determine the probability *P(Buglary, WatsoCalls, RadioNews, ¬Earthquake, ¬Alarm).*

$P(b, w, r, \neg e, \neg a) = P(b) * P(\neg e) * P(\neg a|\neg e, b) * P(r|\neg e) * P(w|\neg a)$

$P(b, w, r, \neg e, \neg a)$ =*0.1*0.999*0.02*0.03*0.1=**0.000005994**=5.994*10^-6*

(b) Determine the probability of "Buglary" given "WatsoCalls" and "RadioNews" are true. "Earthquake", "Alarm" are unknown

$$P(B|r, w) = \alpha * \sum_e \sum_a P(B) * P(e) * P(a|e, B) * P(r|e) * P(w|a)$$

$$P(B|r, w) = \alpha * P(B) * \sum_e (P(e) * P(r|e) \sum_a P(a|e, B) * P(w|a))$$

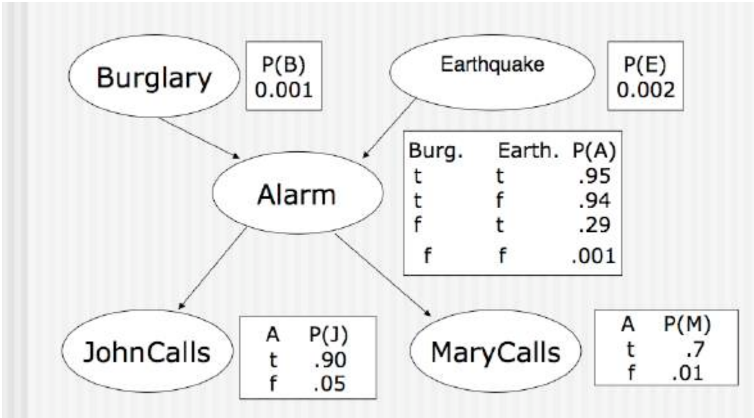
$P(B|r, w) = \alpha * P(B) *($
$(P(e) * P(r|e) * \left(P(a|e, B) * P(w|a) + P(\neg a|e, B) * P(w|\neg a)\right) +$
$(P(\neg e) * P(r|\neg e) * \left(P(a|\neg e, B) * P(w|a) + P(\neg a|\neg e, B) * P(w|\neg a)\right))$

$P(b|r, w) = \alpha * 0.1 *$
((0.001*0.99*(0.99*0.95+0.01*0.1))+(0.999*0.03*(0.98*0.95+0.02*0.1)))= α *0.0028894

$P(\neg b|r, w) = \alpha * 0.9 *$
((0.001*0.99*(0.5*0.95+0.5*0.1))+(0.999*0.03*(0.1*0.95+0.9*0.1)))= α *0.00545778

$P(b|r, w)$ =0.0028894/(0.0028894+0.00545778)= 0.346153

2. Belief Networks



Determine the probability of "MaryCalls" if "Burglary", "Erthquake" are false,"Alarm","JoghnCalls" and are unknown

P(M|not b, not e, A)

$$P(M|\neg b, \neg e, A) = \alpha \sum_{A} P(\neg b)P(\neg e)P(A|\neg b, \neg e)P(M|A)$$

$$P(m|\neg b, \neg e, A) = \alpha(P(\neg b)P(\neg e)(P(a|\neg b, \neg e)P(m|a) + P(\neg b)P(\neg e)(P(\neg a|\neg b, \neg e)P(m|\neg a))$$

$$P(m|\neg b, \neg e, A) = \alpha P(\neg b)P(\neg e)((P(a|\neg b, \neg e)P(m|a) + (P(\neg a|\neg b, \neg e)P(m|\neg a))$$

and

$$P(\neg m|\neg b, \neg e, A) = \alpha P(\neg b)P(\neg e)((P(a|\neg b, \neg e)P(\neg m|a) + (P(\neg a|\neg b, \neg e)P(\neg m|\neg a))$$

$$P(m|\neg b, \neg e, A) = \alpha * 0.999 * 0.998 * (0.001 * 0.7 + 0.999 * 0.01) = \alpha * 0.0107$$

and

$$P(\neg m|\neg b, \neg e, A) = \alpha * 0.999 * 0.998 * (0.001 * 0.3 + 0.999 * 0.99) = \alpha * 0.9863$$

$$P(m|\neg b, \neg e, A) = \frac{0.0107}{0.0107 + 0.9863} = 0.0107$$

$$P(\neg m|\neg b, \neg e, A) = \frac{0.9863}{0.0107 + 0.9863} = 0.9893$$