Solutions to Exercises for Module 5 – Defining Bayesian Network Structure

Exercise 1: Create a Bayesian network with two Boolean nodes A and B. Make B the child of A and define the NPTs respectively as:

NPT for node A:

True	False
80	20

NPT for node B:

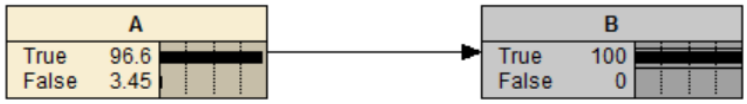
A	True	False
True	70	30
False	10	90

Construct an 'equivalent' version of the model in which A is the child of B.

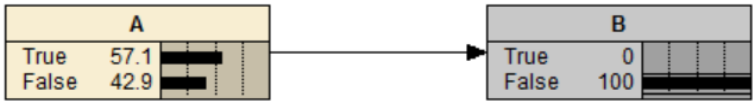
Solution:

The NPT for B has to be the marginal probabilities for B that you see in the first model (i.e. 58% and 42% for true and false respectively).

To calculate the necessary NPT values for A, such as the probability for A being True when B is True, you simply run the first model with evidence selection for B to determine the probability of A. For example, the probability for A being True when B is True is:



And the probability of A being True when B is False is:

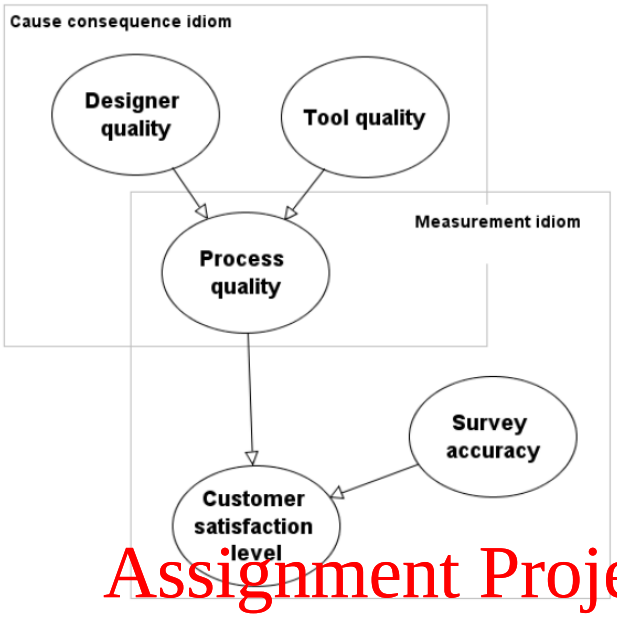


Therefore the NPT for node A is:

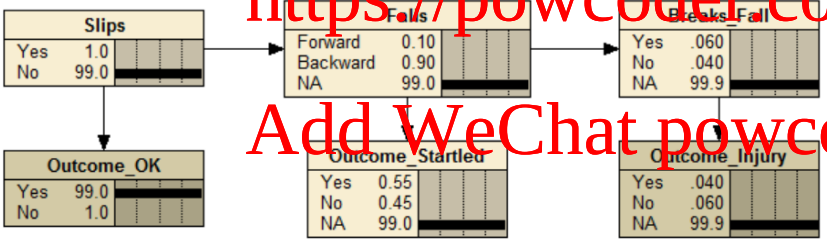
B1	True	False
True	96.6	3.4
False	57.1	42.9

Exercise 2: Use the cause-consequence idiom and the measurement idiom to model the following uncertain scenario: “Using good designers and good tools increases the quality of a product and surveys show that improved product quality increases customer satisfaction”.

Solution:



Exercise 3: Explain why the following Bayesian network has 39 NPT entries in total.



Solution:

Node ‘Slips’ has 2 states and no parents, therefore has an NPT with 2 entries

Node ‘Outcome OK’ has 2 states with one parent with 2 states, therefore has an NPT with 2 x 2 = 4 entries

Node ‘Falls’ has 3 states with one parent with 2 states, therefore has an NPT with 3 x 2 = 6 entries

Node ‘Outcome Startled’ has 3 states with one parent with 3 states, therefore has an NPT with 3 x 3 = 9 entries

Node ‘Breaks Fall’ has 3 states with one parent with 3 states, therefore has an NPT with 3 x 3 = 9 entries

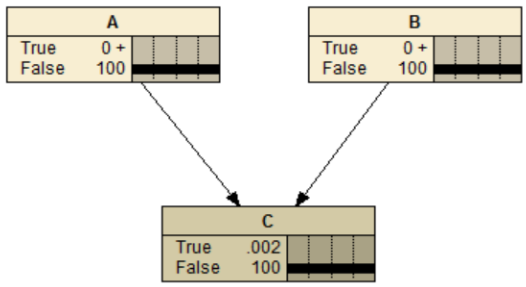
Node ‘Outcome Injury’ has 3 states with one parent with 3 states, therefore has an NPT with 3 x 3 = 9 entries

Total NPT entries = 2 + 4 + 6 + 9 + 9 + 9 = 39

Exercise 4: A, B and C are three rare medical conditions. A and B both have an incidence of about one in 1,000 people. In a large sample of 600,000 patients it is discovered that every patient having either condition A or B also had condition C, and patients that had neither condition A or B did not have condition C. From prior knowledge obtained separately to the large sample of patients, it is known that there are no instances of a patient having condition C if they have both condition A and B. Construct a Bayesian network containing A, B and C. Complete the NPTs for the network using logic and then change the NPTs to what they would be if you had learned them from the data for 600,000 patients.

Solution:

The Bayesian network constructed from logic:



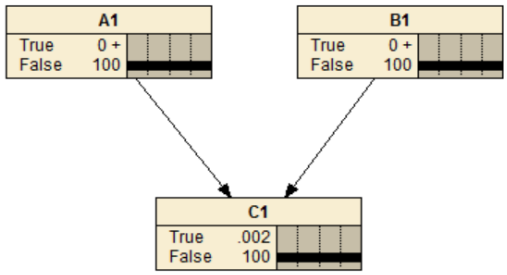
NPT for nodes A and B:

True	False
1.00e-3	99.999

NPT for node C:

A	B	True	False
True	True	0	100
True	False	100	0
False	True	100	0
False	False	0	100

The Bayesian network constructed from the data for 600,000 patients:



NPT for nodes A1 and B1:

True	False
1.00e-3	99.999

NPT for node C1:

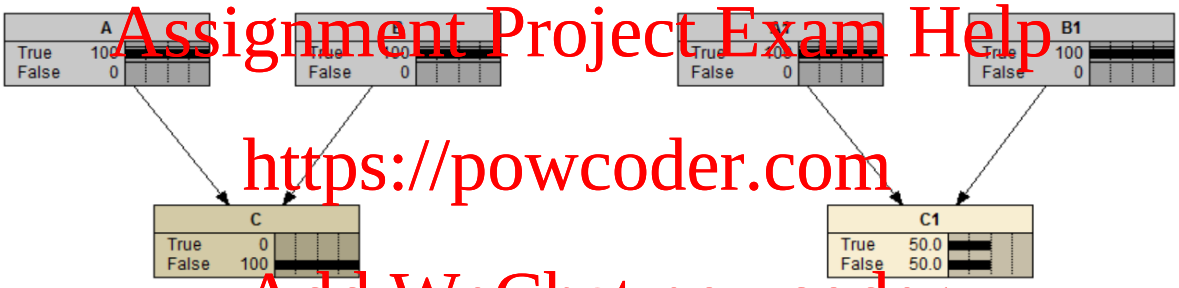
A1	B1	True	False
True	True	50	50
True	False	100	0
False	True	100	0
False	False	0	100

The difference between the two models are the probabilities for the scenario where both A and B are True, that is, those patients that have conditions A and B. From logic we know that a patient who has both conditions A and B cannot have condition C. However, in a data set for 600,000 patients, there are no records of a patient having both conditions A and B as well as condition C. Therefore, if the NPTs were learned from the data, the probability of C being True or False if A and B are both True would be unknown, that is, in a state of maximum uncertainty. In a state of maximum uncertainty the probability of C being True or False if A and B are both True would be equal, that is, 50% chance of being True and 50% chance of being False.

The difference between the models when logic is used to complete the NPTs versus when data is used causes the models to give different predictions for patients that have both conditions A and B. The model constructed from logic gives the correct prediction.

Model from logic

Model from data

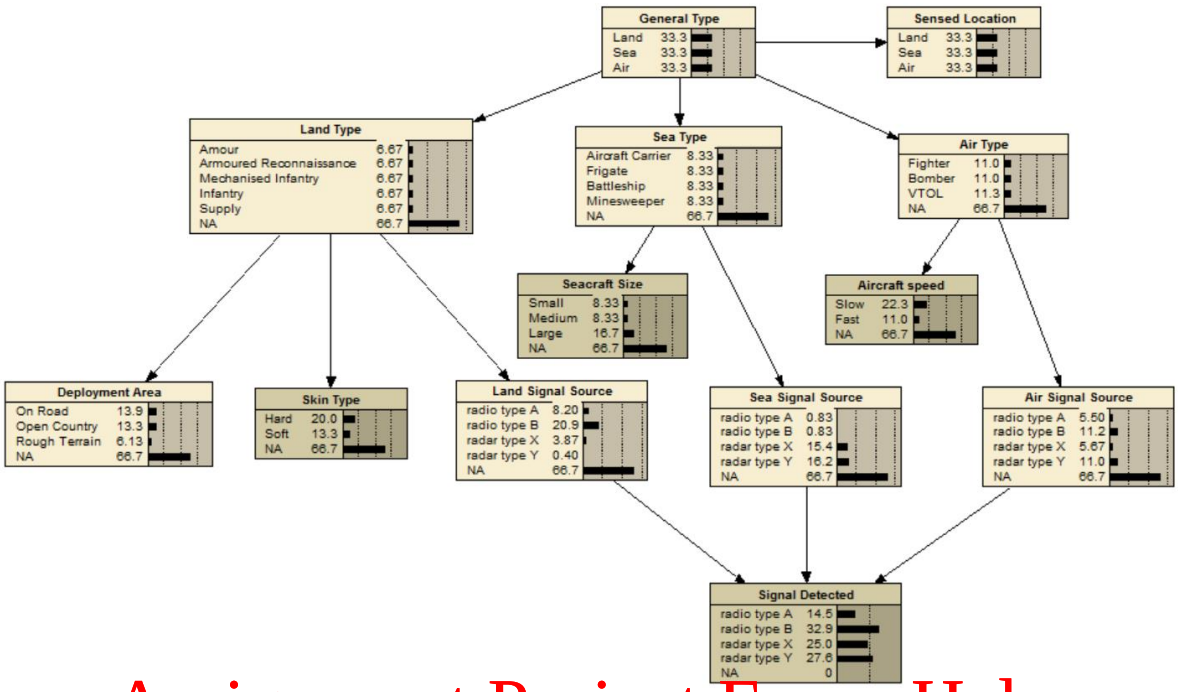


The moral of this story is:

- Sometimes you have to trust logic to provide a far more informed quantitative judgement than you will get from data alone.
- Even really big datasets can be insufficient for small problems.
- Trusting logic can save you a whole lot of unnecessary data collection.

Exercise 5: You have been employed by the military to develop a system for identifying enemy military craft based on the type of signal they emit and where the signal is detected. Construct a Bayesian network that will help you to predict the most likely enemy craft depending on where it is detected (Land, Sea or Air) and the type of signal that it is using (radio type A, radio type B, radar type X and radar type Y).

Solution:



If you detect an enemy craft signal over land and the signal is radar type X, what would be the most likely type of craft?

Solution:

The most likely type of craft is Mechanised Infantry.