14.8 Consider the network for car diagnosis shown in Figure 14.21.
a. Extend the network with the Boolean variables IcyWeather and StarterMotor .
b. Give reasonable conditional probability tables for all the nodes.
c. How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming that no conditional independence relations are known to hold among them?
d. How many independent probability values do your network tables contain?
e. The conditional distribution for Starts could be described as a noisy-AND distribution. Define this family in general and relate it to the noisy-OR distribution.

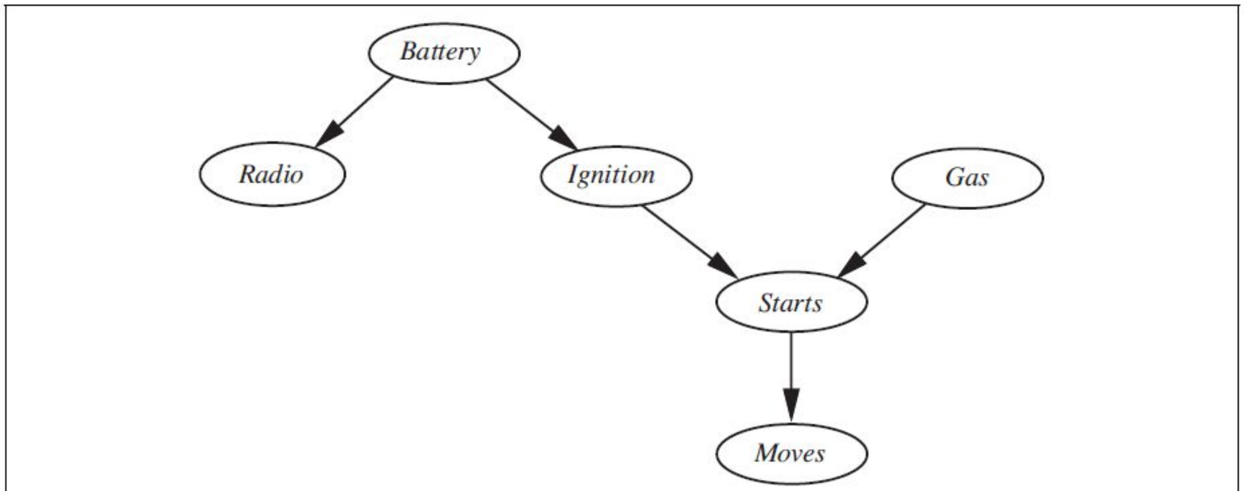


Figure 14.21 A Bayesian network describing some features of a car's electrical system and engine. Each variable is Boolean, and the *true* value indicates that the corresponding aspect of the vehicle is in working order.

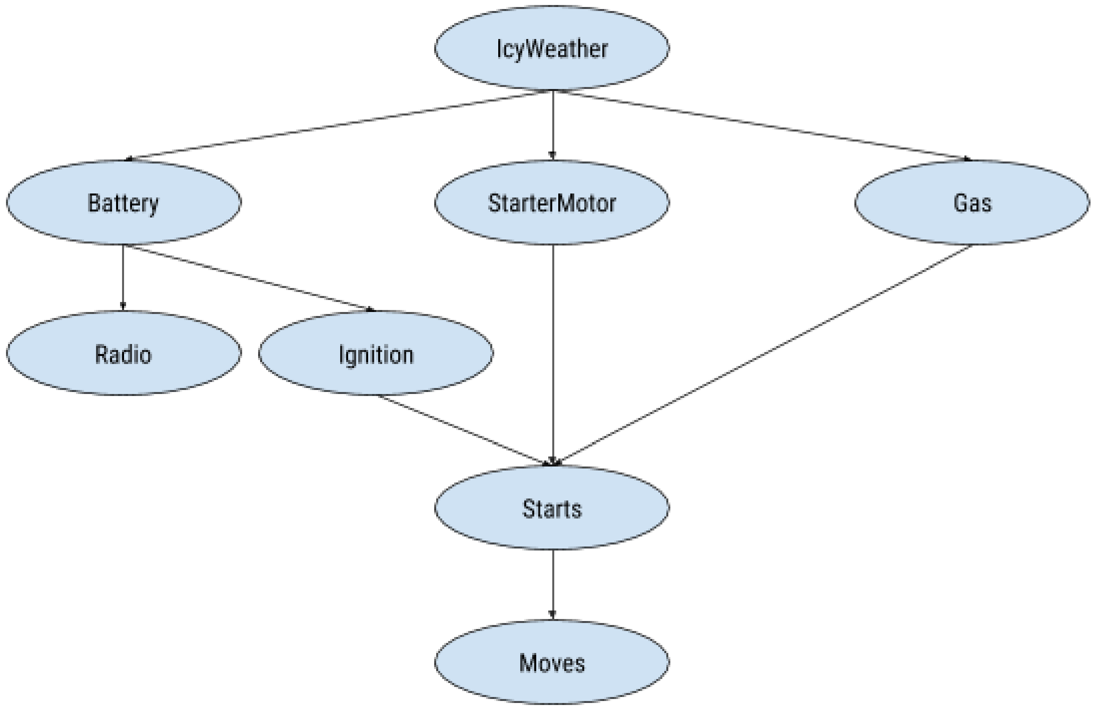
a. Extend the network with the Boolean variables IcyWeather and StarterMotor .

A starter (also self-starter, cranking motor, or starter motor) is a device used to rotate (crank) an internal-combustion engine so as to initiate the engine's operation under its own power. Starters can be electric, pneumatic, or hydraulic. In the case of very large engines, the starter can even be another internal-combustion engine. A battery is a can full of chemicals that produce electrons. The chemical reactions inside of batteries take place more slowly when the battery is cold, so the battery produces fewer electrons. The starter motor therefore has less energy to work with when it tries to start the engine, and this causes the engine to crank slowly. The cold weather could freeze the StarterMotor. Gasoline, like any other liquid, evaporates less when it is cold. You have seen this -- if you pour water onto a hot sidewalk it will evaporate a lot faster than it will from a cooler place like a shady sidewalk. When it gets really cold, gasoline evaporates slowly so it is harder to burn it (the gasoline must be vaporized to burn).

So one possible conclusions about the description above:

- *IcyWeather* is not an effect of any car device. It must be boolean variable with no parents at all.
- *Battery* can be affected directly by *IcyWeather*.
- *StarterMotor* can be affected directly by *IcyWeather* and is required to start the motor.
- *Gas* is affected directly by IcyWeather (*)

(*) Here *Gas=False* could mean that there is no *Gas* or that *Gas* is not functional because of its state.



b. Give reasonable conditional probability tables for all the nodes.

It is logical to expect that *IcyWeather* relies on geographical local weather.

IcyWeather
P(IcyWeather=True)
0.1

Battery fails to work 5% of the time when *IcyWeather* is True.

Battery	
IcyWeather	*P(Battery=True | ...)*
True	0.95
False	0.999

The *StarterMotor* gets frozen 1% of the time when *IcyWeather* is True and it fails 1 time in 100,000 when weather is Ok.

StarterMotor	
IcyWeather	*P(StarterMotor=True | ...)*
True	0.99
False	0.99999

The *Gas* get unusable 1 time in 10,000 times when *IcyWeather* is True and get unusable 1 in a million when Weather is Ok. May be because its age.

Gas	
IcyWeather	*P(Gas=True | ...)*
True	0.9999
False	0.999999

The *Radio* fails 1 in 100,000 when *Battery* works and it does not work when *Battery* if False.

Radio	
Battery	*P(Radio=True | ...)*
True	0.99999
False	0.0

The *Ignition* fails 1 in thousand (probably because human handling error) when *Battery* is True and it does not work when *Battery* is False. Here I ignore the possibility that ignition could works with no battery at all if somebody helps by pushing the car.

Ignition	
Battery	*P(Ignition=True | ...)*
True	0.999
False	0.0

The car could start just when *Ignition*, *StarterMotor* and *Gas* are True.

Starts			
Ignition	*StarterMotor*	*Gas*	*P(Starts=True | ...)*
True	True	True	0.99999
True	True	False	0
True	False	True	0
True	False	False	0
False	True	True	0
False	True	False	0
False	False	True	0
False	False	False	0

The car can be moved only if car has started. But in some occasion could happen that car was started just to close the window or to keep the radio on.

Moves	
Starts	*P(Moves=True | ...)*
True	0.999
False	0.0

c. How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming that no conditional independence relations are known to hold among them?

For Boolean variables, once you know that the probability of a true value is p, the probability of false must be $1-p$, so we often omit the second number, as in Figure 14.2. In general, a table for a Boolean variable with k Boolean parents contains 2^k independently specifiable probabilities. A node with no parents has only one row, representing the prior probabilities of each possible value of the variable. [AIMA 3E, p.512]

Lets calculate the maximum number of parents for every node in a Bayesian Network with 8 Boolean variables:

Node 1: 0 parents, CTP Size: $2^0 = 1$
Node 2: 1 parents, CPT Size: $2^1 = 2$

Node 3: 2 parents, CPT Size: $2^2 = 4$
Node 4: 3 parents, CPT Size: $2^3 = 8$
Node 5: 4 parents, CPT Size: $2^4 = 16$
Node 6: 5 parents, CPT Size: $2^5 = 32$
Node 7: 6 parents, CPT Size: $2^6 = 64$
Node 8: 7 parents , CPT Size: $2^7 = 128$

Sum(CPTs) = 1+2+4+8+16+32+64+128 = $2^8 - 1$ = 255

d. How many independent probability values do your network tables contain?

Observing the number of parent links that every node in the network presents we can calculate it:
$2^0 + 2^1 + 2^1 + 2^1 + 2^1 + 2^1 + 2^3 + 2^1 = 1 + 2 + 2 + 2 + 2 + 2 + 8 + 2 = 21$

(*) Or we can count the number of rows in the above CPTs

e. The conditional distribution for Starts could be described as a noisy-AND distribution. Define this family in general and relate it to the noisy-OR distribution.

We say that the conditional distribution for *Starts* could be described as a *noisy-AND* distribution because the only situation where the car *Starts* is when all of its parents are True (*AND(parents(Starts))==True*). In my example I said that *P(Starts | AND(parents(Starts))=True) = 0.99999* because it could be possible that other unexpected condition happens that makes the car no starting. We can add an extra node named *Leak* that will cover all other unknown variables in the world that are required to be True to make the car starts. Then, the new CPT for *Starts* node will present four given variables: *Ignition, StarterMotor, Gas* and *Leak,* and when all of them are True, the probability that the car starts would be exactly 1 (one).

In the other hand, the idea with noisy-OR is very similar:

noisy-OR
The noisy-OR model allows for uncertainty about the ability of each parent to cause the child to be true (the causal relationship between parent and child may be inhibited). The model makes two assumptions. First, it assumes that all the possible causes are listed. (If some are missing, we can always add a so-called leak node that covers "miscellaneous causes."). Second, it assumes that inhibition of each parent is independent of inhibition of any other parents.[AIMA 3E, p.518-519]

With this assumption:
$P(\neg x_i \mid all\ variables\ in\ parents(X_i)\ are\ False) = 0$

And if we count with all probabilities q_j where:
$q_j = P(\neg x_i \mid all\ variables\ in\ parents(X_i)\ are\ False\ with\ exception\ of\ X_j\ that\ is\ True)$

Then, from this information above and the noisy-OR assumptions, the entire CPT can be built. The general rule is that:

$$P(x_i \mid parents(X_i)) = 1 - \prod_{\{j: X_j = True\}} q_j$$