Solutions to Selected Exercises

Chapter 1, Exercise 3

Simpson's paradox warns us that the conclusion is NO. In fact this data comes from a real study. Inspecting the number of patients in each of the four groups it becomes clear that exactly the opposite was true Treatment B was more effective):

	Small stones	Large stones	Overall successes	Overall success rate
Treatment A	81/87	192/263	273/350	78%
Treatment B	243/270	55/80	289/350	83%

Chapter 1, Exercise 5

This is an example of what is normally referred to as the 'two envelopes problem' and it is discussed extensively on wikepedia and elsewhere. If there is no maximum possible prize value then it can be argued that the rational decision is always to switch boxes. If the box you choose contains \$100 then there is an evens chance the other box contains \$50 and an evens chance it contains \$200. If you do not switch you have won \$100. If you do switch you are just as likely to decrease the amount you win as increase it. However, if you win the amount increases by \$100 and if you lose it only decreases by \$50. So your expected gain is *positive* (rather than neutral). In fact, if this was repeated 10 times you would expect about 5 times to get \$200 and 5 times to get \$50 making a total of \$1250 with an average win of \$125. This is 25% more than if you never switched. The formal 'proof' of this is as follows:

Suppose that the amount in the first box is *S*. Then there is a probability ½ the other box contains 2*S* and a probability ½ it contains ½*S*. The expected amount[1] in the other box is therefore:

$$\frac{1}{2}\times 2S+\frac{1}{2}\times\frac{1}{2}S=\frac{5}{4}S$$

i.e. 25% more than S

The result also means that, if the boxes are not opened and you switch to the second box then you should switch back to the first box if given a second option to switch. In fact you should continue to switch if given the option.

However, it has been argued that there are problems with the above 'proof', primarily because it assumes the largest prize is infinite (with thanks to reader Hugh Panton for pointing this out in an earlier version of these solutions). In fact, it seems reasonable to assume that there is a finite maximum prize, even if we allow that maximum to be as large as we like. With this assumption it turns out that we can prove (without dispute) that ***there is no difference to your expected winnings*** if you stick or switch.

Without loss of generality we can assume the possible prizes are $1,2,4,8,\ldots,2^n$

[1] The expected value of an uncertain variable will be defined formally in Chapter 4

So 2^n is the maximum prize for some n. Note that if the box contains the 2^n prize then you MUST lose 2^{n-1} if you switch.

For example, if n=3 then the possible prizes are 1,2,4,8 and the possible pairings are (1,2),(2,4),(4,8). Crucially, note that the 'middle' numbers 2 and 4 appear twice whereas the 'end numbers' 1 and 8 appear only once. So, the probability of getting an end number is 1/2n (1/6 in this case) and the probability of getting a middle number is 1/n (1/3 in this case).

Now, we will show that the expected increased 'gain' from switching must be 0.

To see this we first deal with the 'end points' 1 and 2^n.

If the number is 1 then switching must get us a 2 and so the 'gain' is 1. Since there is a 1/(2n) probability of a 1, the expected gain in this case is

$$\frac{1}{2n} \times 1 = \frac{1}{2n}$$

However, if the number is 2^n then switching must get us a 2^{n-1} and so the 'gain' is actually LOSS of 2^{n-1}. Since there is a 1/(2n) probability of a 2^n, the expected gain in this case is

$$-\left(\frac{1}{2n} \times 2^{n-1}\right) = -\left(\frac{2^{n-1}}{2n}\right)$$

So the total expected gain from the two end points is actually a LOSS of

$$\frac{2^{n-1}}{2n} - \frac{1}{2n}$$

$$= \frac{1}{2n}(2^{n-1} - 1) \qquad (1)$$

So, for example, when n=3 the expected gain from '1' is 1/6 but the expected loss from '8' is 4/6 making a total LOSS of 3/6 = ½.

We will show this is the same as the expected GAIN from the 'middle' numbers:

For the number 2 there is ½ probability that we switch to a 1 (and hence lose 1) and a ½ probability we switch to a 4 (and hence gain 2). Since there is a 1/n chance of getting a 2, the expected gain is:

$$\frac{1}{n} \times \frac{1}{2}(-1 + 2) = \frac{1}{2n}(1)$$

For the number 4 there is ½ probability we switch to a 2 (and hence lose 2) and a ½ probability we switch to a 8 (and hence gain 4). Since there is a 1/n chance of getting a 4 the expected gain is:

$$\frac{1}{n} \times \frac{1}{2}(-2+4) = \frac{1}{2n}(2)$$

Note that if n=3 then the total expected gain from the middle number is

$$\frac{1}{6} + \frac{2}{6} = \frac{1}{2}$$

which is exactly the same as the expected loss from the end numbers.

In general, for the kth middle number the expected gain is

$$\frac{1}{2n}(k)$$

So the total expected gain from the middle numbers is:

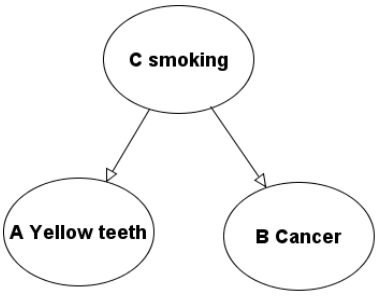
$$\frac{1}{2n}(1) + \frac{1}{2n}(2) + \frac{1}{2n}(4) + \cdots \frac{1}{2n}(2^{n-1})$$

$$= \frac{1}{2n}(1 + 2 + 4 + .. + 2^{n-1})$$

$$= \frac{1}{2n}(2^{n-1} - 1)$$

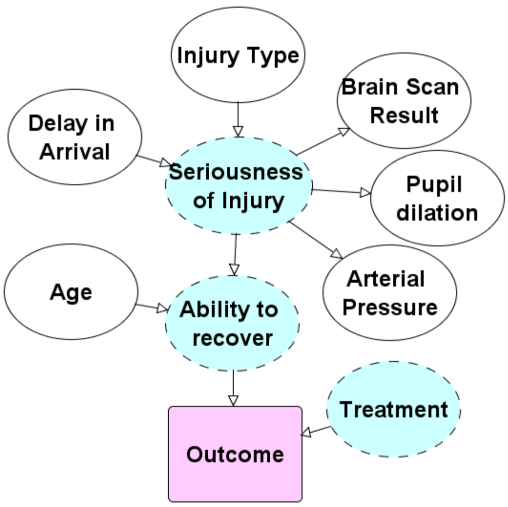
And this is exactly the same as the expected LOSS (equation (1)) of the end numbers.

Chapter 2, Exercise 1



Chapter 2, Exercise 5

The model does not take account of actual treatment that did take place for the patients and which could take place for future patients; nor does it take account of what might have happened if the treatment was different (this is called a counterfactual problem). It also fails to recognize the difference between causal factors affecting the seriousness' of injury (such as delay in arrival) compared with those that result from its measurement (such as pupil dilation). It will therefore be unable to predict accurately from profile data which patients most urgently need treatment and how the chance of survival is reduced if treatment is delayed. A more suitable causal model (that addresses all of these concerns by introducing the necessary hidden interventions and explanatory factors and distinguishing between cause and effect) is:



Chapter 3, Exercise 7

(a) Each of the 5 students must have a different number. For each group of 5 students with the same coloured jumper there are 4 colours to choose from. Choose the numbers for the students in C(6,5) =6 ways and choose the colours for the students in C(4,1)=4 ways. By the product rule, choose the group in 24 ways.

(b) The 2 students must share the number but belong to different groups. Choose the number in C(6,1) = 6 ways and choose the colour in C(4,2) =6 ways.

By the product rule, choose the 2 students in 36 ways. The remaining 3 students have numbers chosen from the remaining 5 numbers. Each of these students can be chosen from any of the 4 groups, possibly the same. Choose the number in C(5,3) = 10 ways; choose the colour in C(4,1) x C(4,1) x C(4,1) = 64 ways. By the product rule, choose the 3 students in 640 ways. By the product rule, choose the group in 36 x 640 ways.

Chapter 3, Exercise 13

2/3, 1/3 and 0.

Chapter 4, Exercise 5

Each of these scenarios has a different probability as follows:

1. In a family with exactly three children the probability they each have the same birthday is approximately 1/133225. This is indeed approximately equal to 7.5 in a million as stated (although curiously when we asked a number of people to tell us what they understood by the statement "the odds are ... 7.5 in a million" most people thought it meant 7.5 million to one, which is very different).

   The explanation is quite straightforward. If we assume all three birthdates are independent then the probability that the second child has the same birthday as the first is 1/365. That's because whatever that birthday happens to be (29 Jan in this case) that day is just as likely to be the birthday of the second child as any of the other 364 days of the year.

   Similarly, the probability that the third child has this birthday is also 1/365. So the probability that all three have this birthday is 1/365 times 1/365 which is equal to 1/133225.

   In practice the probability will be higher because parents are more likely to conceive at certain times of the year and so the probability that the second child's birthday is the same as the first is greater than 1/365. As an extreme example imagine a couple who only make love between May and September. Then any of their children will almost certainly have birthdays between February and June and so the probability of the second child's birthday being the same as the first is more like 1/151.

2. In a family of more than three children the probability of exactly three having the same birthday is much higher. For example if there are four children (a,b,c, and d), then we can consider four different combinations of three children (a,b,c), (a,b,d), (a,c,d) and (b,c,d). For each of these four combinations the probability of all three having the same birthday is 1/133225. So the probability that at least one of the combinations has the same birthday is 4/133225, i.e. it is four times more likely. With five children there are TEN different combinations of three children so the probability is ten times greater etc.

The final question we need to ask is: is this story newsworthy? The answer is no. For every million families involving at least three children we would EXPECT there to be at least 8 families in which three children share the same birthday. In the UK there are certainly more than a million families involving at least three children. **It would therefore have been far more newsworthy if it was found that NO family in the UK contained three children with the same birthday.**

Chapter 4, Exercise 6

For a continuous probability distribution any exact 'point' value (such as 169.2 in this case) is always 0 because the 'area under the curve' for any point is 0. It is therefore only meaningful to consider probabilities of ***ranges*** of values. In this case we could set a small (but non-zero) 'error margin' *e* (for example *e*=0.2) and calculate the probability between 169.2 plus or minus *e* (it is approximately 0.008 in this case).

Chapter 5, Exercise 4

Let T be event "Person is terrorist". Then we know P(T)=1/100 assuming a person is sampled at random from the room.

Let D be event "Lie detector says person tested is a terrorist". Then we assume the statement about "95% accuracy" is interpreted as:

P(D|T) = 0.95, P(not D| T) = 0.05

P(not D | not T) = 0.95, P(D | not T) = 0.05

What we have to calculate is P(T|D):

$$P(T \mid D) = \frac{P(D \mid T) * P(T)}{P(D \mid T) * P(T) + P(D \mid \overline{T}) * P(\overline{T})} = \frac{0.95 * 0.01}{0.95 * 0.01 + 0.05 * 0.99} = \frac{0.0095}{0.059} = 0.1627$$

If we also assume that people are selected and tested randomly until the first positive ID is made then the above probability gives us the required answer (this is the probability that the first person who tests positive is actually the terrorist).

Chapter 5, Exercise 6

The answer is 1/3, not ½ as most people assume. There are three possibilities: (B, G), (G, B), (G, G) where (B, G) means the first child born was a boy and the second child born was a girl etc. Assuming equal probability of B and G, each of these scenarios is equally likely (i.e. 1/3).

Chapter 5, Exercise 7 ("Rule of 5")

Let *X* be the (unknown) population median for number of minutes spent in a car yesterday. We have to find the probability that *X* lies between the lowest and highest value from a sample of 5 people. But this probability is the same as one minus the probability that *X* is either bigger than ALL of the 5 samples numbers or is smaller than ALL of the 5 samples numbers. Each of these probabilities is 1/32 (see below) so the probability that either is true is 1/16. Hence the probability that *X* lies between the lowest and highest value is 15/16, i.e. 93.75%.

The reason for the 1/32 probability is as follows: For any randomly selected person we know that there is a 50% probability that person spent longer than *X* minutes in a car yesterday. The probability that 5 randomly selected people all spent longer than *X* minutes is therefore

$$\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

(Similarly for the probability that 5 randomly selected people spent less than *X* minutes).

The ramifications of this result (discussed in the book "How to Measure Anything" by Douglas Hubbard) is that our uncertainty about completely unknown things can actually be radically minimized by very small scale sampling. Hence his assertion that you really can measure anything. If you were to sample 10 rather than 5 people then (by the same arguments as above) there is a better than 99.8% probability (511/512) that the (unknown) population mean lies between the lowest and highest value in your sample.

Chapter 5, Exercise 8

$P(E \mid H_p) = 25/36$ (assuming the probability of rolling a 6 on the 'fixed' die is 5/6)
$P(E \mid H_d) = 1/36$ (assuming the probability of rolling a 6 on a fair die is 1/6)
LR=25 (the prosecution likelihood is 25 times greater than the defence likelihood) so whatever the prior odds of guilt are they increase by a factor of 25 after observing the evidence.

Chapter 5, Exercise 10

Since the LR is 1 we know that $P(E \mid H_p) = P(E \mid H_d)$

$$P(H_P \mid E) = \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)P(H_P) + P(E \mid H_D)P(H_D)}$$

$$= \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)P(H_P) + P(E \mid H_P)P(H_D)} \text{ since } P(E \mid H_P) = P(E \mid H_D)$$

$$= \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)(P(H_P) + P(H_D))} \text{ since } P(E \mid H_P) = P(E \mid H_D)$$

$$= \frac{P(E \mid H_P)P(H_P)}{P(E \mid H_P)} \text{ since } P(H_P) + P(H_D) = 1 \text{ as } P(H_P), P(H_D) \text{ mutually exclusive and exhaustive}$$

$$= P(H_P)$$

Chapter 5, Exercise 11

We have:

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

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

$$P(E|H_p, H_d) = 1$$

$$P(E|H_p, \text{not } H_d) = 1/6$$
$$P(E|\text{not } H_p, H_d) = 1/6$$

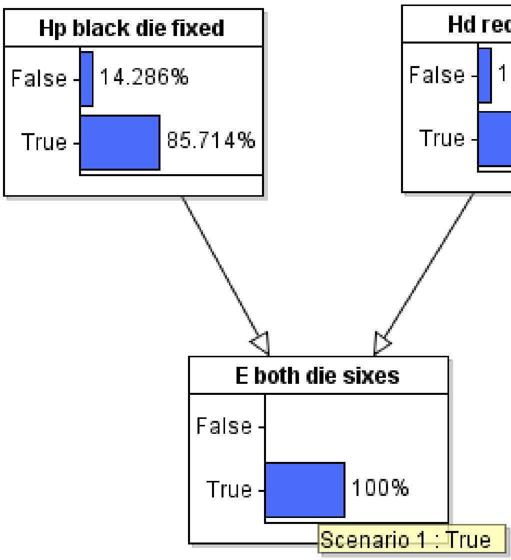
$P(E|\text{not } H_p, \text{not } H_d) = 1/36$

It follows that P(E|H_p)=P(E|H_d) = 7/12 because:

$$P(E|H_p) = P(E|H_p, H_d)P(H_d) + P(E|H_p, \text{not } H_d)P(\text{not } H_d)$$
$$= \frac{1}{6} + \frac{1}{6} \times \frac{1}{2} = \frac{7}{12}$$

$$P(E|H_d) = P(E|H_d, H_p)P(H_p) + P(E|H_d, \text{not } H_p)P(\text{not } H_p)$$
$$= \frac{1}{6} + \frac{1}{6} \times \frac{1}{2} = \frac{7}{12}$$

So LR=1, but the evidence is ***not*** neutral as can be seen from the results of running the model here:



Evidence not neutral

The fact that $P(H_p|E)$= 6/7 = 0.85714 tells us that the prosecution hypothesis is now very likely. Just because the defence hypothesis has increased by the same amount is essentially irrelevant.
Formally, the calculations are based on Bayes' theorem and noting that the marginal $P(E)$ is

$$P(E) = P(E|H_p, H_d)P(H_p)P(H_d) + P(E|H_p, \text{not } H_d)P(H_p)P(\text{not } H_d) + P(E|\text{not } H_p, H_d)P(\text{not } H_p)P(H_d) + P(E|\text{not } H_d, \text{not } H_p)P(\text{not } H_p)P(\text{not } H_d)$$

$$= \left(1 \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{6} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{6} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{36} \times \frac{1}{2} \times \frac{1}{2}\right)$$

=49/144=0.34028

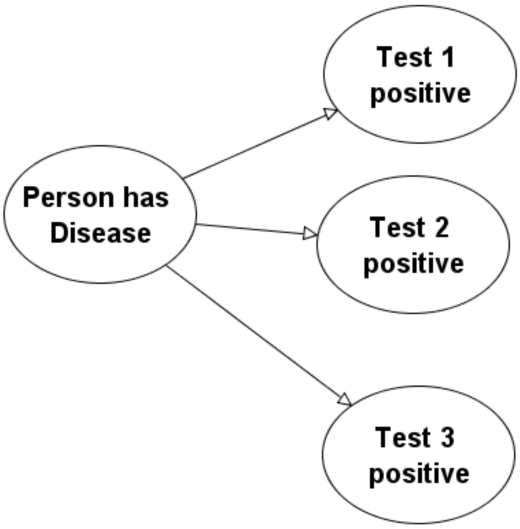
Hence by Bayes'

$$P(H_p) = \frac{P(E|H_p) \times P(H_p)}{P(E)} = \frac{\frac{7}{12} \times \frac{1}{2}}{\frac{49}{144}} = \frac{6}{7}$$

Thus, the fact that the posterior for H_p and H_d increase in the same proportions from their priors is less important than the fact that the posterior for H_p is now more likely than unlikely.

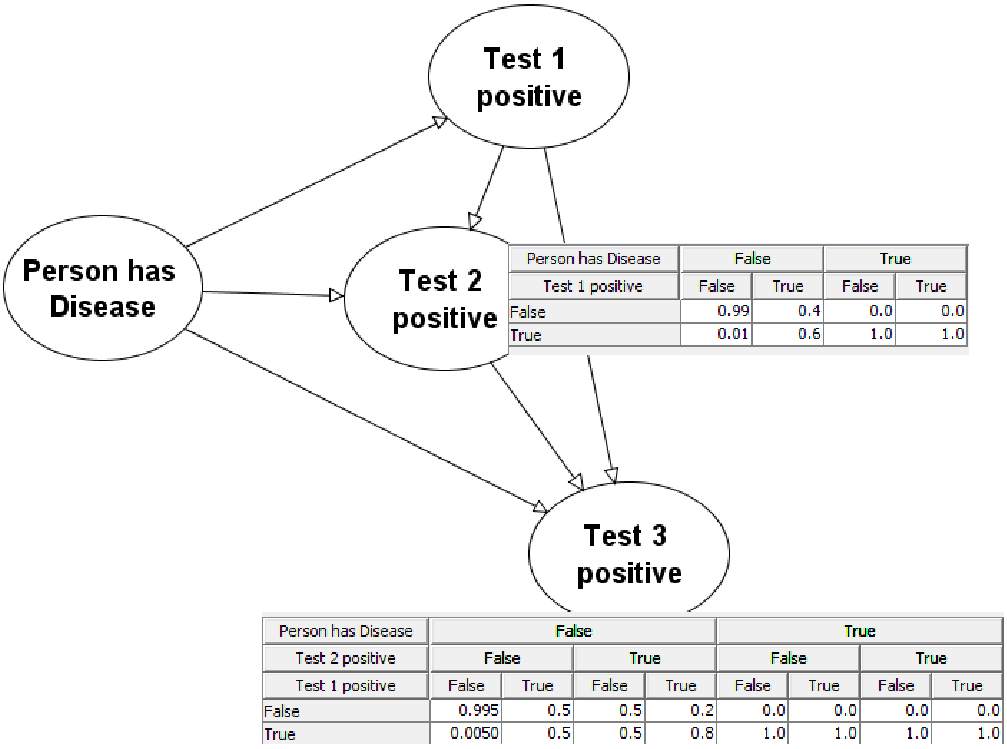
Chapter 6, Exercise 1

Assuming independent tests:



The NPTs for Test2 and Test3 should be identical to the NPT for Test 1. When Test 1 is positive the model calculates the probability the person has disease is 1.964%. When both tests 1 and 2 are positive this probability rises to 28.592%; and when all three tests are positive it rises to 88.899%.

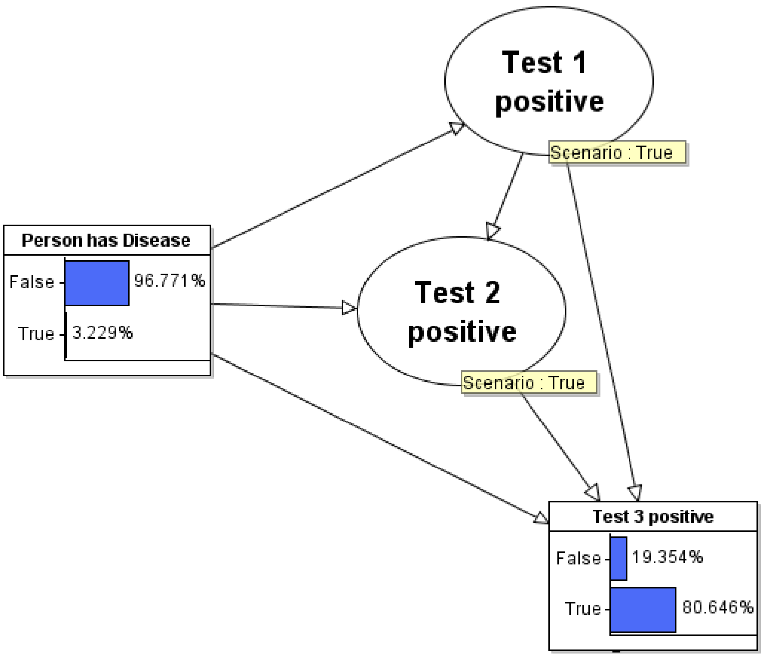
However, suppose the tests are not independent. Then the model should be revised to:



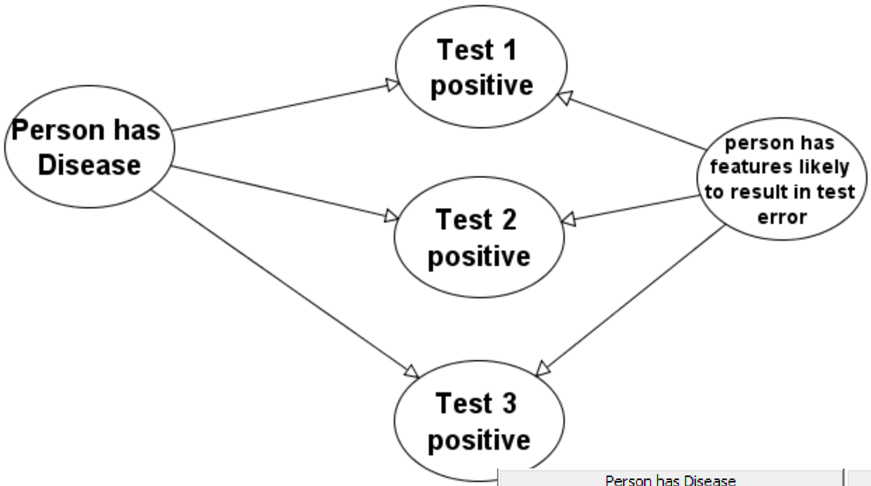
Person has Disease	False		True	
Test 1 positive	False	True	False	True
False	0.99	0.4	0.0	0.0
True	0.01	0.6	1.0	1.0

Person has Disease	False				True			
Test 2 positive	False		True		False		True	
Test 1 positive	False	True	False	True	False	True	False	True
False	0.995	0.5	0.5	0.2	0.0	0.0	0.0	0.0
True	0.0050	0.5	0.5	0.8	1.0	1.0	1.0	1.0

The NPTs defined here assume fairly strong dependence between the tests (note also the NPTs have to deal by default with 'impossible' state combinations such as Disease = True and Test 1 Positive = False). With these assumptions we get:
Tests 1 and 2 both positive: Probability of disease only increases to 3.229%
All 3 tests positive: Probability of disease only increases to 4.002%.
In fact, when we run the model with the first two tests positive, the model ***predicts*** that the third test will be positive with probability 80.464%. So when it does in fact test positive it provides little new support:



Another way that the tests can be dependent is if the person being tested has features likely to result in test error:



Person has Disease	False		True	
person has features likely to result in test error	False	True	False	True
False	0.95	0.9	0.0	0.0
True	0.05	0.1	1.0	1.0

In the above version we have used the same NPT for each test.
In this case we get:
Tests 1 and 2 both positive: Probability of disease only increases to 13.805%
All 3 tests positive: Probability of disease only increases to 64.023%.

Chapter 6, Exercise 2 (ii)

P(A, B, C, D, E, F) = P(A)P(B)P(C|A,B)P(D|B)P(E|B,C,D)P(F|C,D)

Chapter 6, Exercise 6

i. 90%, 1%, 9.1%
ii. 50%, 8.33%, 5.5%
iii. The important point is that ‘soft evidence’ will generally ***not*** be equal to the posterior probability. The only reason it is in the case of the “Smoker” node is because the prior probability for the states (yes/no) of the “Smoker” node were equal, so when we enter soft evidence in that node the posterior probabilities become equal to the soft evidence values. However, in the “Visit to Asia” the prior probability for the states are not equal; in fact the prior probability of ‘yes’ is so low that the “90%” soft evidence can only shift posterior from 1% to 8.33%. You have to be very confident to make a major shift (if you enter soft evidence of 99.9% ‘yes’ then the posterior moves to 91% ‘yes’),

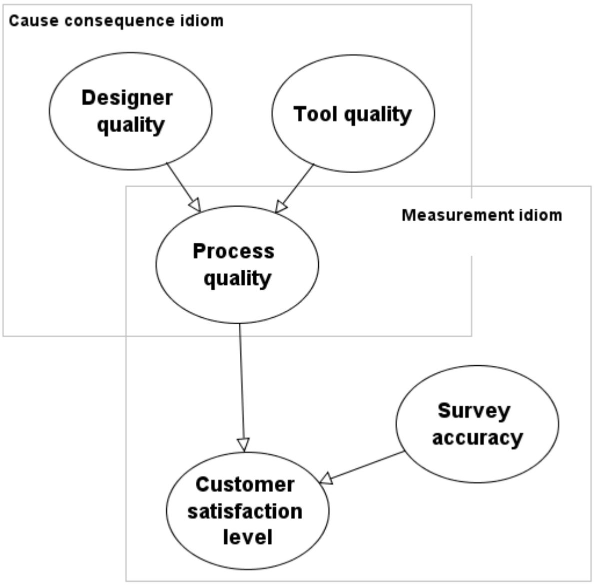
Chapter 7, Exercise 1

The models are identical even though they have the “opposite” direction for “causality”.

Chapter 7, Exercise 2

The NPT for B has to be the marginal probabilities for B that you see in the first model (i.e. 0.42 and 0.58 for false and true respectively. To calculate the necessary NPT values for A - such as the value for A being True when B is true, you simply run the first model with the corresponding observation (B true).

Chapter 7, Exercise 3



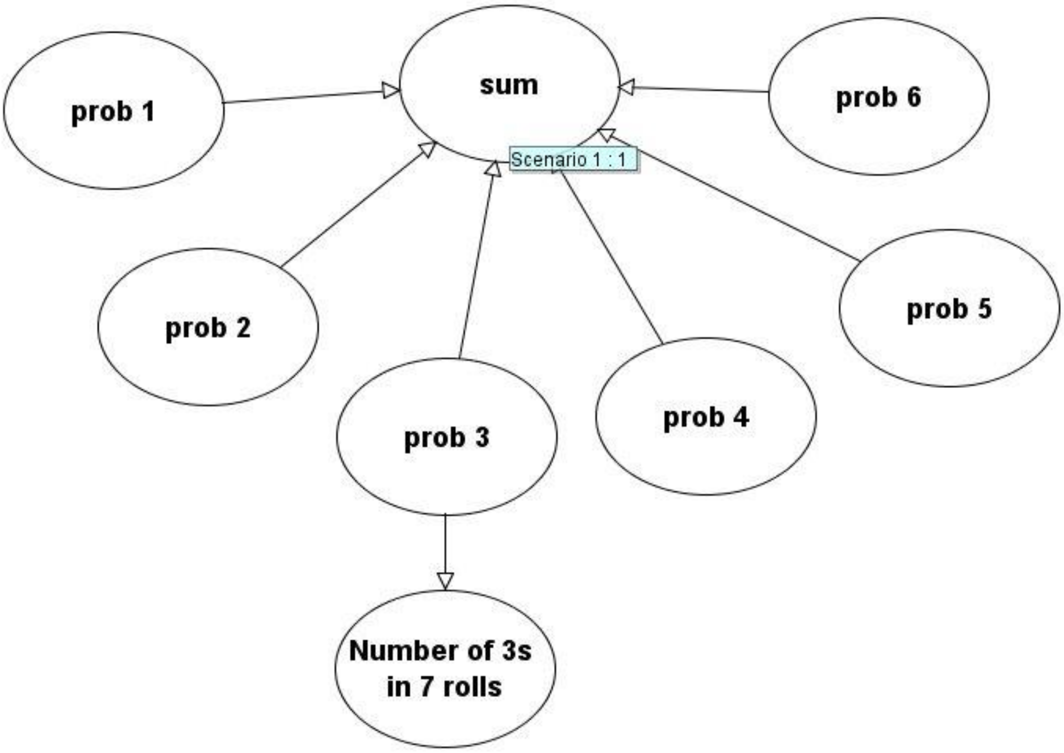
Chapter 8, Exercise 1

5^4=625
The shear size of the table makes it impractical to complete manually, but even if it was attempted the most serious problem is ensuring consistency. For example, suppose all nodes have states (poor, below average, average, above average, high) and that the parents of A are nodes B, C, D. Now suppose that the entry for A being 'below average' given that B, C, D are all average is 0.2. Then, if B, C, D are all expected to have a positive effect on A you would have to make sure that whenever B, C, D are all at least 'average', the entry for A being 'poor' is at least 0.2.
To avoid the problems of manually completing such a table you could, of course, reduce the number of states in each node (but even 3 states for each results in an NPT that is extremely difficult to complete manually). Hence the best solution is to use a predefined function (such as a weighted mean) for the NPT.

Chapter 9, Exercise 2

With 10 iterations the mean is 30.094, median 29.992, variance 111.73.
With 100 iteration the mean is 30, median 30, variance 100.06.
So as the number of iterations increases the result gets very close to the 'true' distribution. But even at a low number of iterations the accuracy is reasonable.

Chapter 9, Exercise 4
The basic BN we need is this one:



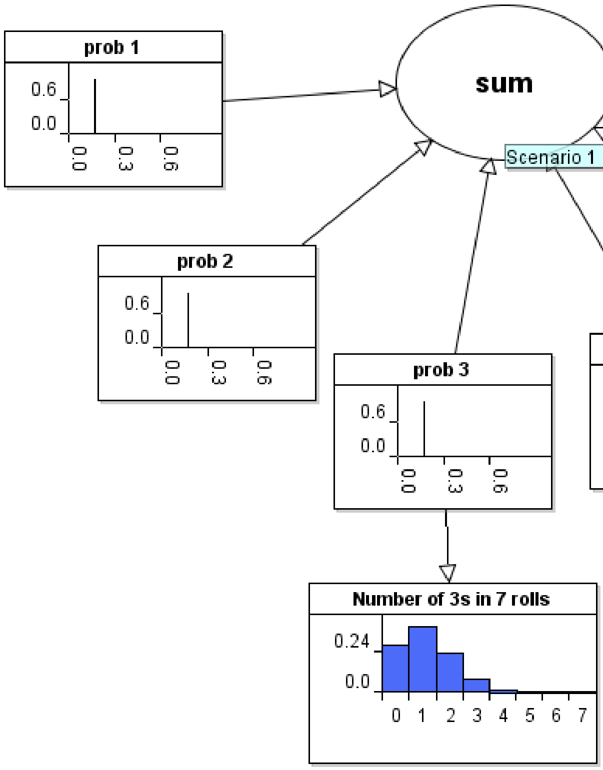
Here the nodes prob 1, prob 2, ..., prob 6 represent the probability of rolling the number 1,2,...,6 respectively.

The node 'sum' is a logical constraint on the model (it is the sum of the six probabilities and, because of the probability axioms this sum must be 1 assuming that no outcome other than 1,2,...,6 is possible from rolling the die).

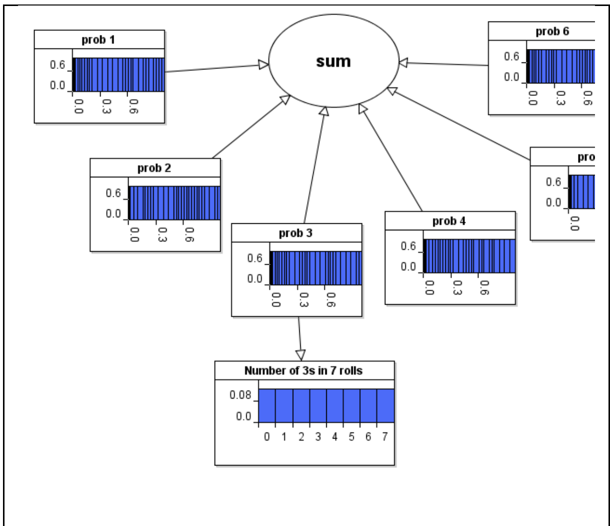
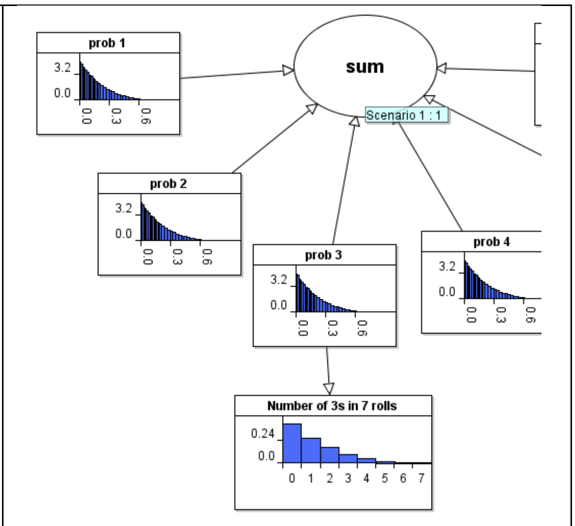
The node "number of 3's in 7 rolls" is defined as a Binomial distribution (the probability that the number is 7 is simply $p3^7$ where p3 is the probability of getting a 3).

In the model we have to set some prior probability distribution on each of the nodes prob 1, prob 2, ..., prob 6 (the particular choice of prior is what distinguishes a,b, and c). Before we enter the evidence of the 7 rolls of 3, the model - when calculated - displays the prior marginal probabilities. Thus:

a) The probability of each Pk is ***exactly*** 1/6, so the prior probability distributions Pk looks like this:

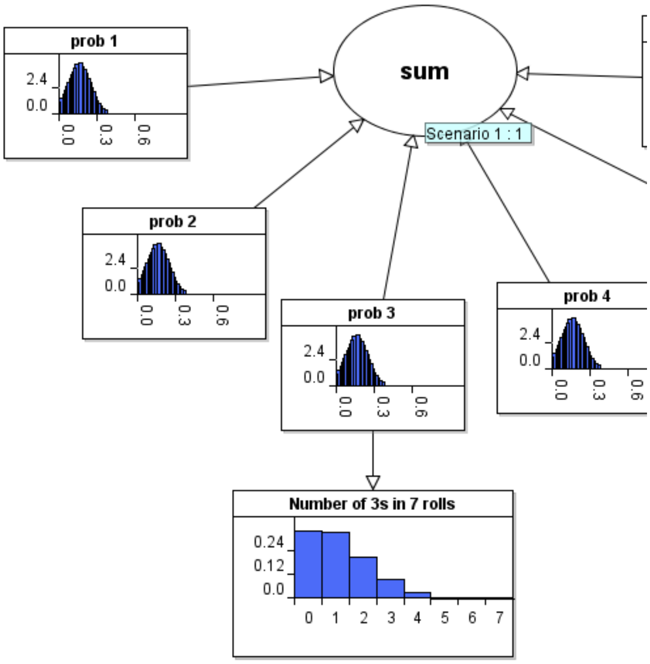


b) In the absence of any prior knowledge of the die, the probability distribution for each Pk is ***uniform*** over the interval 0-1 (meaning any value is just as likely as any other), so the prior probability distribution for each node Pk looks like this (before and after we enter the sum=1 constraint):

	
Before sum constraint	after sum constraint

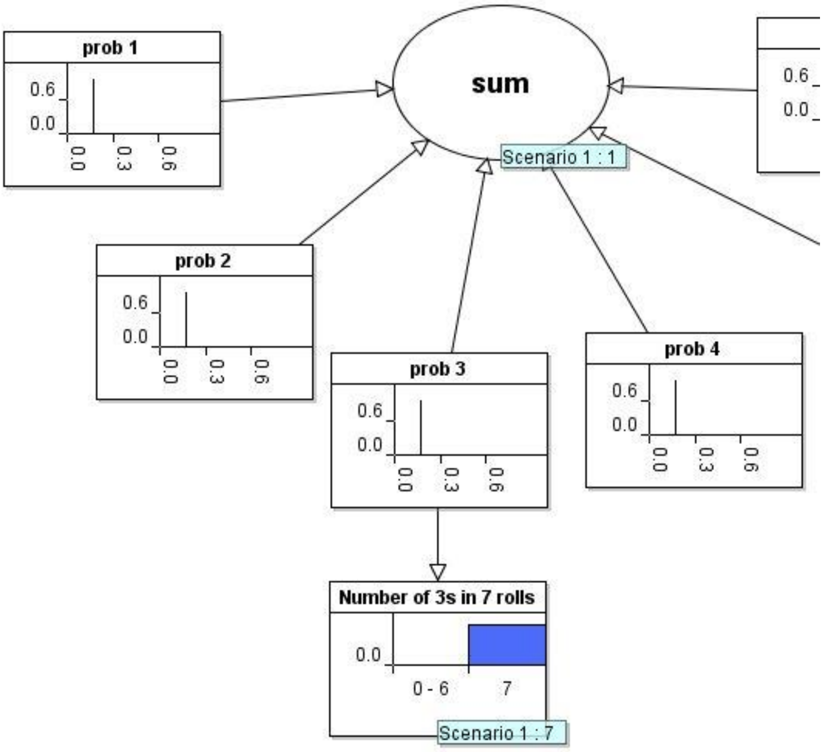
c) Here it seems reasonable to specify the prior distribution for Pk to be a narrow bell curve[2] centred on 1/6:

[2] We have used a Truncated Normal distribution with mean 1/6 and variance 0.01 over the range 0-1



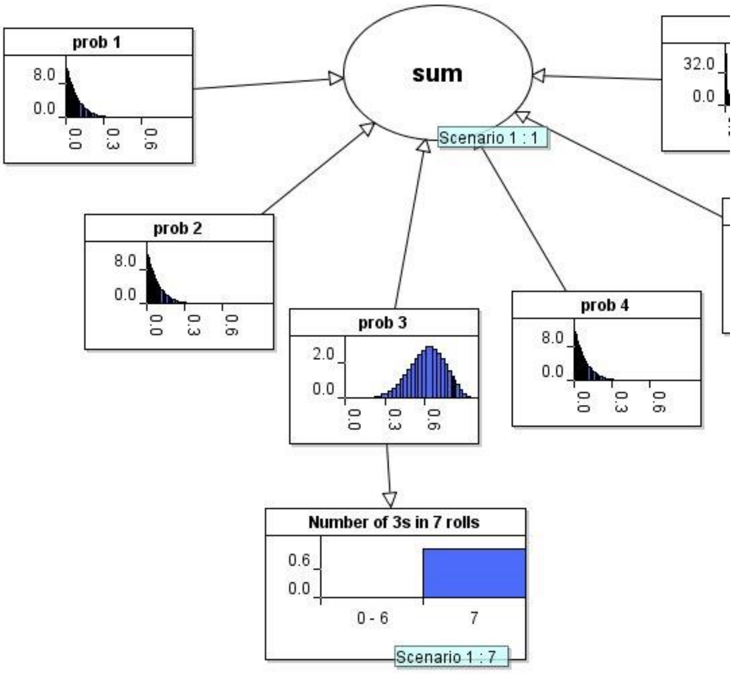
When we enter the evidence of seven 3's in 7 rolls, the Bayesian calculations (performed here using AgenaRisk) result in an updated posterior distribution for each of the nodes prob 1, ..., prob 6:

in a) the posterior for each node is unchanged:



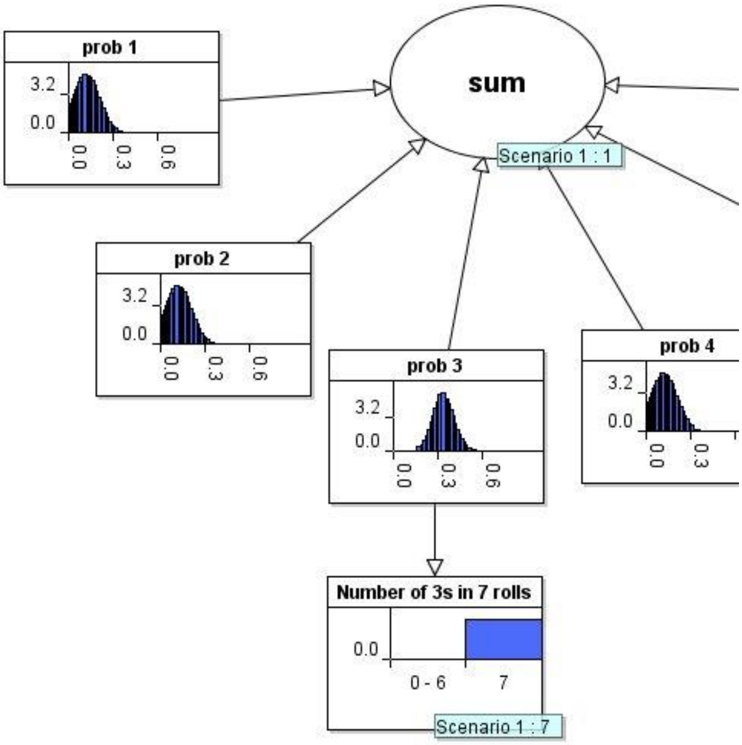
i.e. the probability the next roll of the die will be 1,2,3,4,5,6 are all respectively still 1/6.

in b) the posteriors are:



The prob 3 is now a distribution with mean 0.618. The other probs are all reduced accordingly to distributions with mean about 0.079. So in this case the probability of rolling a 3 next time is about 0.618 whereas each of the other numbers has a probability about 0.079

In c) the posteriors are:



The prob 3 is now a distribution with mean 0.33. The other probs are all reduced accordingly to distributions with mean about 0.13. So in this case the probability of rolling a 3 next time is about 0.33 whereas each of the other numbers each has a probability about 0.13.

And what about the statistician? Well a classical statistician cannot give any prior distributions so the above approach does not work for him. What he might do is propose a 'null' hypothesis that the die is 'fair' and use the observed data to accept or reject this hypothesis at some arbitrary 'p-value' (he would reject the null hypothesis in this case at the standard p=0.01 value). But that does not provide much help in answering the question. He could try a straight frequency approach in which case the probability of a three is 1 (since we observed 7 out of 7 threes) and the probability of any other number is 0.

Chapter 10, Exercise 2

The classical hypothesis test involves a null hypothesis 'no weight loss' and a predefined p-value such as 0.01. This means we would reject the null hypothesis if the probability of no weight loss given the data is less than 1%. The model shows that the probability of no weight loss for the drug Precision is certainly less than 1% (it is the probability that 'precision less than 0' is true). So we would certainly reject the null hypothesis for Precision. But for the drug Oomph, the probability of no weight loss is more than 1%. Hence we cannot reject the null hypothesis for Oomph at the 0.01 level. So in 'classical' terms Precision is somehow more acceptable as a weight loss drug than Oomph. Yet, the (Bayesian) "Hypothesis" node clearly shows that weight loss with Oomph is greater than Precision 93% of the time.

Chapter 12, Exercise 2

The most likely 'explanation' is 'very low' operational usage (58%).

Chapter 13, Exercise 1

If we assume that the 1/1000 DNA match probability is truly representative, that there are no other links of the defendant to the crime, and that the DNA collection, analysis and testing were perfectly accurate, then we can reasonably conclude that $P(E| H_d)=1/1000$.
If we assume that the DNA collection, analysis and testing were perfectly accurate, then we can reasonably conclude that $P(E \mid H_p) = 1$. Hence the LR is 1000.
The assumptions are generally unrealistic because there will be uncertainty about whether the DNA sample tested was the one collected at the scene, and whether there was cross contamination at any point in the process; also it is known that DNA testing is not 'prefect'.
However, the biggest concern is that it is impossible to define $P(E \mid H_p)$ and $P(E \mid H_d)$ meaningfully without knowing something about the priors $P(H_p)$, $P(H_d)$ (in strict Bayes' terms[3] we say the likelihoods and the priors are all ***conditioned on some background knowledge K***). For example, suppose the DNA trace was found on the murder victim. Now consider two extreme values that may be considered appropriate for the prior $P(H_p)$, derived from different scenarios used to determine K :

a) $P(H_p) = 0.5$, where the defendant is one of two people seen grappling with the victim before one of them killed the victim;
b) $P(H_p) = 1/40$ million.where nothing is known about the defendant other than he is one of 40 million adults in the UK who could have potentially committed the crime.

[3] Specifically, the priors $P(H_p)$, $P(H_d)$, really refer to $P(H_p|K)$ and $P(H_d|K)$ respectively. The likelihoods must take account of the same background knowledge K that is implicit in these priors. So the 'real' likelihoods we need are $P(E|H_p, K)$ and $P(E|H_d, K)$.

Whereas a value for P(E | H_d) = 1/1000 seems reasonable in case b), it is clearly not in case a). In case a) the defendant's DNA is very likely to be on the victim irrespective of whether or not he is guilty. This suggests a value of P(E | H_d) close to 1. It follows that, without an understanding about the priors and the background knowledge, we can end up with vastly different LRs associated with the same hypotheses and evidence.

Chapter 13, Exercise 2

The model needed is shown in Figure 1 (for the first part you can ignore the evidence of motive) with the prior probabilities shown:

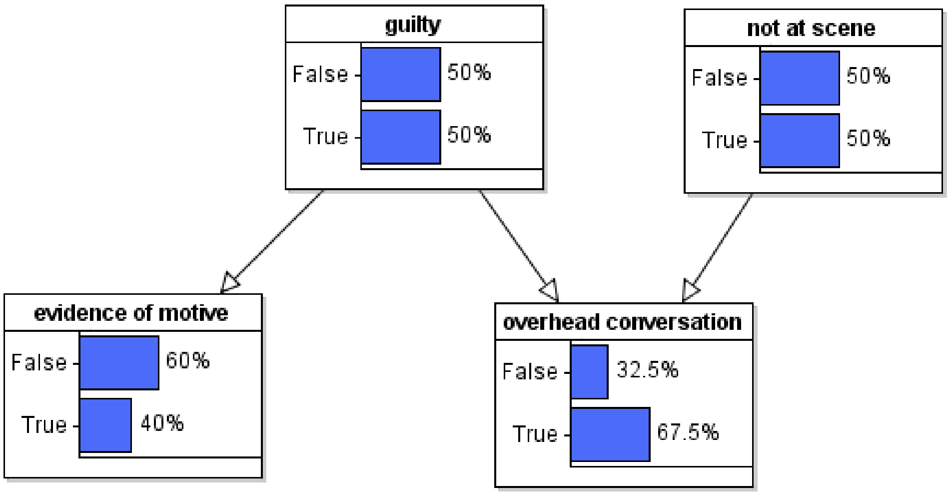


Figure 1 Prior probabilities for murder model

When you enter the evidence of the overheard conversation you will see that both 'guilty' and 'not at scene' increase to 0.666.
Figure 2 shows the results of observing the motive evidence E'. In this case P(H_p | E') = 0.875, while P(H_d | E') is unchanged at 0.5.

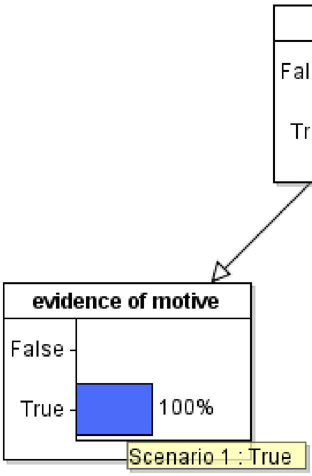


Figure 2 Posterior probabilities for murder model after observing evidence of motive

When we now observe *E* (Figure 3) we see that the probability of H_p, that is P(H_p |E', E), jumps to 0.933. The evidence *E* therefore may be sufficient in this case to convince a jury to convict (if there were, say a threshold of 90% certainty required).

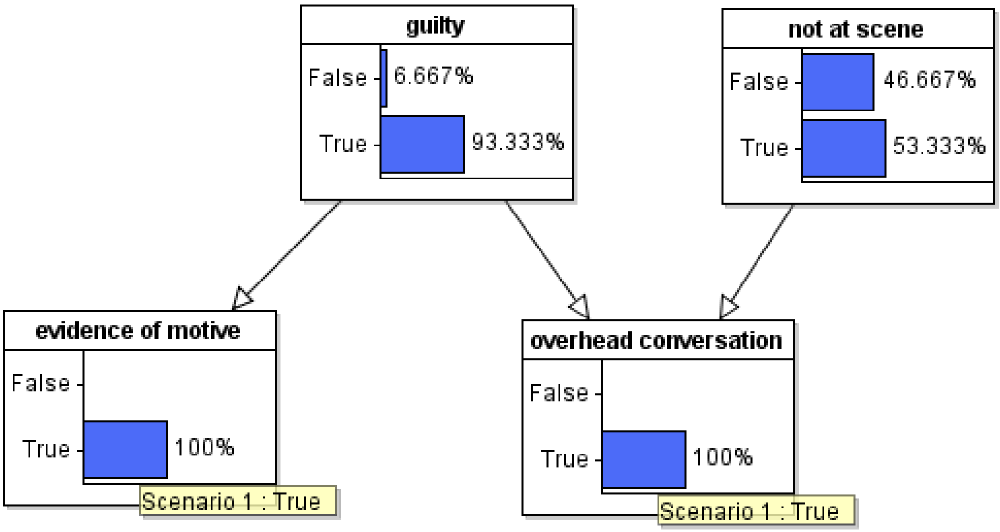


Figure 3 Posterior probabilities for murder model after observing evidence of motive and overheard conversation

Chapter 13, Exercise 4

The solution to this is provided in the model "Exercise 13_4_forensics_with_errors.cmp" on www.bayesianrisk.com