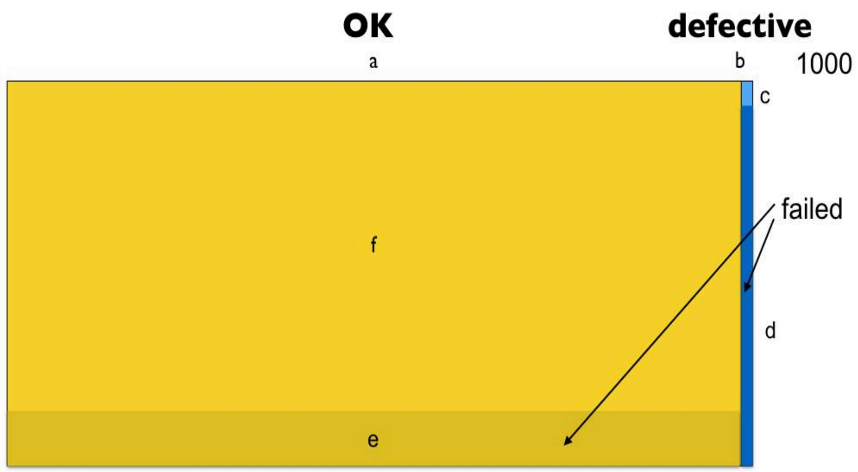
Optimisation and Statistical Data Analysis

Exercise Set 8 Solutions

Problem 1

Part a



- "3% produced are defective" : $b = 30$
- $a + b = 1000 \Rightarrow a = 970$
- "fail 90% of defective" : $d = 0.9 \times b = 27$
- $c + d = b \Rightarrow c = 3$
- "fail 20% of OK" : $e = 0.2 \times a = 194$
- $e + f = a \Rightarrow f = 776$

Probability that a failed bottle is defective = $\dfrac{d}{d+e} = \dfrac{27}{27+194} = \dfrac{27}{221} \approx 0.1222$

Probability that a passed bottle is defective = $\dfrac{c}{c+f} = \dfrac{3}{3+776} = \dfrac{3}{779} \approx 0.0039$

Part b

- "3% produced are defective" : : $\text{Prob}(\text{def}) = 0.03$, $\text{Prob}(\neg\text{def}) = 0.97$
- "fail 90% of defective" : $\text{Prob}(\text{fail} \mid \text{def}) = 0.9$
- "fail 20% of nondefective" : $\text{Prob}(\text{fail} \mid \neg\text{def}) = 0.2$

Conditional probabilities given "fail" :

$$\text{Prob}(\text{def} \mid \text{fail}) \propto \text{Prob}(\text{fail} \mid \text{def})\,\text{Prob}(\text{def}) = 0.9 \times 0.03 = 0.027$$

$$\text{Prob}(\neg\text{def} \mid \text{fail}) \propto \text{Prob}(\text{fail} \mid \neg\text{def})\,\text{Prob}(\neg\text{def}) = 0.2 \times 0.97 = 0.194$$

Total probability formula : $\text{Prob}(\text{fail}) = 0.027 + 0.194 = 0.221$

Probability that a failed bottle is defective = $\text{Prob}(\text{def} \mid \text{fail}) = \dfrac{0.027}{0.221} = 0.1222$

Conditional probabilities given "pass" (i.e. not "fail") :

$$\text{Prob}(\text{def} \mid \neg\text{fail}) \propto \text{Prob}(\neg\text{fail} \mid \text{def})\,\text{Prob}(\text{def}) = (1 - \text{Prob}(\text{fail} \mid \text{def}))\,\text{Prob}(\text{def}) = 0.1 \times 0.03 = 0.003$$

$$\text{Prob}(\neg\text{def} \mid \neg\text{fail}) \propto \text{Prob}(\neg\text{fail} \mid \neg\text{def})\,\text{Prob}(\neg\text{def})(1 - \text{Prob}(\text{fail} \mid \neg\text{def}))\,\text{Prob}(\neg\text{def}) = 0.8 \times 0.97 = 0.776$$

Total probability formula : $\text{Prob}(\neg\text{fail}) = 0.003 + 0.776 = 0.779$

Probability that a passed bottle is defective = $\text{Prob}(\text{def} \mid \neg\text{fail}) = \dfrac{0.003}{0.779} = 0.00385$

Part c

Assume that the two inspections are conditionally independent. We use the previous answer

$$\text{Prob}(\text{def} \mid \neg\text{fail}_1) = 0.00385, \quad \text{Prob}(\neg\text{def} \mid \neg\text{fail}_1) = 1 - 0.00385 = 0.99615$$

as the prior for the second inspection.

The conditional probabilities given $\neg\text{fail}_2$:

$$\text{Prob}(\text{def} \mid \neg\text{fail}_1, \neg\text{fail}_2) \propto \text{Prob}(\neg\text{fail}_2 \mid \text{def})\,\text{Prob}(\text{def} \mid \neg\text{fail}_1) = 0.1 \times 0.00385 = 0.000385$$

$$\text{Prob}(\neg\text{def} \mid \neg\text{fail}_1, \neg\text{fail}_2) \propto \text{Prob}(\neg\text{fail}_2 \mid \neg\text{def})\,\text{Prob}(\neg\text{def} \mid \neg\text{fail}_1) = 0.8 \times 0.99615 = 0.79692$$

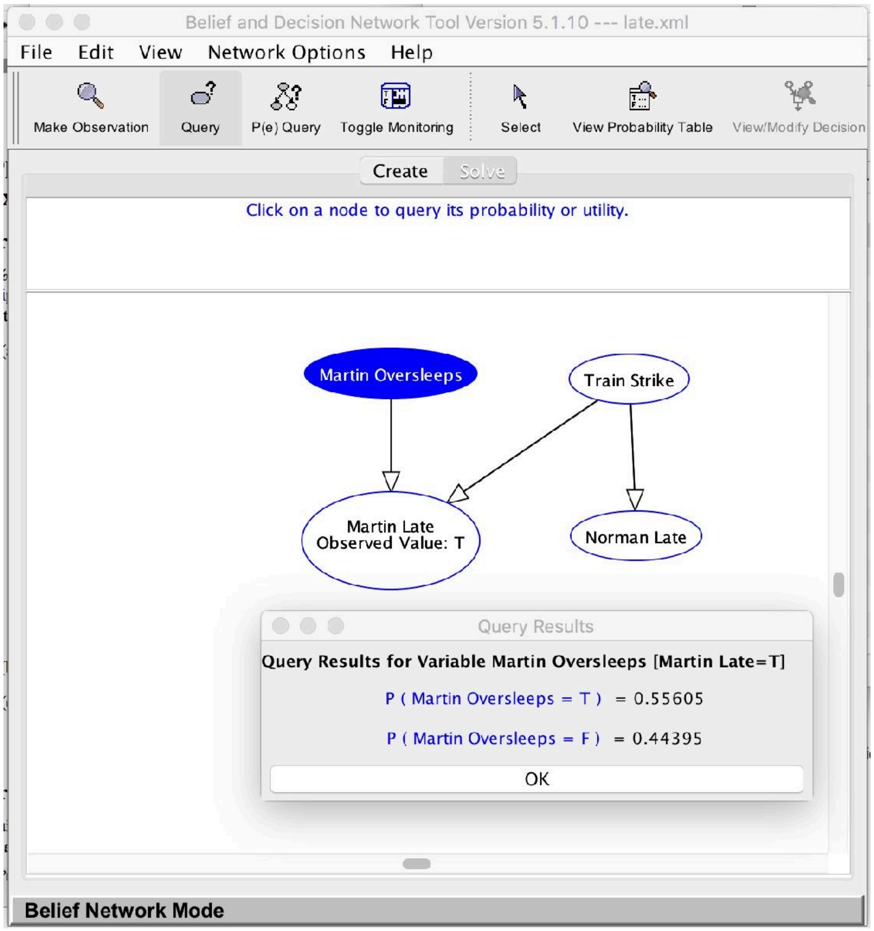
Total probability formula : $\text{Prob}(\neg\text{fail}_2 \mid \neg\text{fail}_1) = 0.000385 + 0.79692 = 0.797305$

Probability that a twice-passed bottle is defective = $\text{Prob}(\text{def} \mid \neg\text{fail}_1, \neg\text{fail}_2) = \dfrac{0.000385}{0.797305} = 0.000483$

Problem 2

Part a

In the *Solve* window, use *Make Observation* to set *Martin Late* to *T* ("true"). Then use *Query* or *Toggle Monitoring* to read the node values. From



we read that

$$\text{Prob}(\text{Martin Oversleeps} \mid \text{Martin Late}) = 0.56$$

Similarly we read

$$\text{Prob}(\text{Train Strike} \mid \text{Martin Late}) = 0.15$$

That is, given this observation, "Martin overslept" is much more probable than "Train strike".

Parts b and c

Set *Martin Late* = *T* and *Norman Late* = *T*. The node values are then

$$\text{Prob}(\text{Martin Oversleeps} \mid \text{Martin Late} \wedge \text{Norman Late}) = 0.51$$

$$\text{Prob}(\text{Train Strike} \mid \text{Martin Late} \wedge \text{Norman Late}) = 0.59$$

That is, given these observations, "Train strike" is slightly more probable than "Martin overslept". Note how Norman's lateness changes our explanation for Martin's lateness.

Using *P(e) Query* we can read the "evidence" for the observations:

$$\text{Prob}(\text{Martin Late} \wedge \text{Norman Late}) = 0.0922$$

Extra problem c

Instructions on how to fill the probability tables of the BN are given at http://staff.utia.cas.cz/vomlel/mh-puzzle.html

Then, with observations First Selected = #1 and Monty opens = #2 , the posterior is

- Prob(Has Prize = #1 | observations) = 0.333,
- Prob(Has Prize = #2 | observations) = 0,
- Prob(Has Prize = #3 | observations) = 0.667,

so the contestant should switch.