Part One:

Question One:

X	Y	Computation	P(X,Y)
0	0	0.3 * 0.3	0.09
0	1	0.3 * 0.7	0.21
1	0	0.7 * 0.8	0.56
1	1	0.7 * 0.2	0.14

I used multiplication of the probabilities from the CPT (Conditional probability table) and Bayes rule.

Question Two:

x -> y -> z

X	Y	Z	Computation	P(X,Y,Z)
0	0	0	0.3 * 0.3 * 0.6	0.054
0	0	1	0.3 * 0.3 * 0.4	0.036
0	1	0	0.3 * 0.7 * 0.8	0.168
0	1	1	0.3 * 0.7 * 0.2	0.042
1	0	0	0.7 * 0.8 * 0.6	0.336
1	0	1	0.7 * 0.8 * 0.4	0.224
1	1	0	0.7 * 0.2 * 0.8	0.112
1	1	1	0.7 * 0.2 * 0.2	0.028

I used the product rule and the conditionally probability tables (CPT) to calculate P(X, Y, Z). We can do this as P(X) * P(Y | X) * P(Z | Y).

Question Three:

i)

P(Z = 0) = 0.054 + 0.168 + 0.336 + 0.112 = 0.670

P(X = 0, Z = 0) = 0.054 + 0.168 = 0.222

ii)

No, X and Z are not truly independent of each other. Z is dependent on Y and Y is depended on X, therefore X cannot be independent of Z. However there is an exception to this: Z is independent from X given Y, however that is a very specific case and doesn't apply all the time- therefore they cannot be independent.

Question Four:

i)

P (X = 1, Y = 0 |Z = 1) = **P(X = 1, Y = 0) / P(Z = 1)**

P(X = 1, Y = 0) = 0.336 + 0.224 = **0.56**

P(Z = 1) = **0.33**

= 0.224 / 0.33

= **0.678**

ii)

P (X = 0 | Y = 0, Z = 0) = **P(X = 0) / P(Y = 0, Z = 0)**

P(X = 0) = **0.3**

P(Y = 0, Z = 0) = 0.054 + 0.336 = **0.39**

= 0.054 / 0.39

= **0.1385**

Part Two: Naive Bayes Method

Question One:

	Spam	**Not Spam**
Feature 1 (True) :	34/51	53/149
Feature 1 (False):	17/51	96/149
Feature 2 (True) :	30/51	86/149
Feature 2 (False):	21/51	63/149
Feature 3 (True) :	23/51	51/149
Feature 3 (False):	28/51	98/149
Feature 4 (True) :	31/51	59/149
Feature 4 (False):	20/51	90/149
Feature 5 (True) :	25/51	50/149
Feature 5 (False):	26/51	99/149
Feature 6 (True) :	18/51	70/149
Feature 6 (False):	33/51	79/149
Feature 7 (True) :	40/51	75/149
Feature 7 (False):	11/51	74/149
Feature 8 (True) :	39/51	52/149
Feature 8 (False):	12/51	97/149
Feature 9 (True) :	17/51	36/149
Feature 9 (False):	34/51	113/149
Feature 10 (True) :	34/51	43/149
Feature 10 (False):	17/51	106/149
Feature 11 (True) :	34/51	87/149
Feature 11 (False):	17/51	62/149
Feature 12 (True) :	40/51	50/149
Feature 12 (False):	11/51	99/149

Question Two:

Probability Spam: 0.649%, Probability Not spam: 99.351%. Probably: Not Spam.

Probability Spam: 57.441%, Probability Not spam: 42.559%. Probably: Spam.

Probability Spam: 59.337%, Probability Not spam: 40.663%. Probably: Spam.

Probability Spam: 0.860%, Probability Not spam: 99.140%. Probably: Not Spam.

```
Probability Spam: 39.097%, Probability Not spam: 60.903%. Probably: Not Spam.
Probability Spam: 55.245%, Probability Not spam: 44.755%. Probably: Spam.
Probability Spam: 1.035%, Probability Not spam: 98.965%. Probably: Not Spam.
Probability Spam: 13.577%, Probability Not spam: 86.423%. Probably: Not Spam.
Probability Spam: 83.465%, Probability Not spam: 16.535%. Probably: Spam.
Probability Spam: 2.882%, Probability Not spam: 97.118%. Probably: Not Spam.
```

Question Three:

Naive Bayes algorithm assumes that the effect of a values predictor of a given class is independent of the values of other predictors, this is called conditional independence. There is a highly likely chance on the spam data that attributes aren't conditionally independent. For example, seeing frequent capital letters, and the text "you won a million dollars" will most likely mean that there will be a fake address. This is not conditionally independent as two attributes give us information on a third.

If two attributes aren't conditionally independent using the Naive Bayes algorithm it would simply multiply them together, ignoring the condition- which will hence give the wrong result. For simplicities sake, lets assume that there are only two attributes that categorise the spam x1 and x2. If they weren't conditionally independent (lets say x2 depends on x1) then Naive Bayes would calculate it to be P(x1, x2) = **p(x1) * p(x2)**. Where as the correct calculation would be P(x1, x2) = **p(x1) * p(x2 | x1)**, as it assumes x1 has occurred when determining p(x2).

Part Three: Bayesian Networks

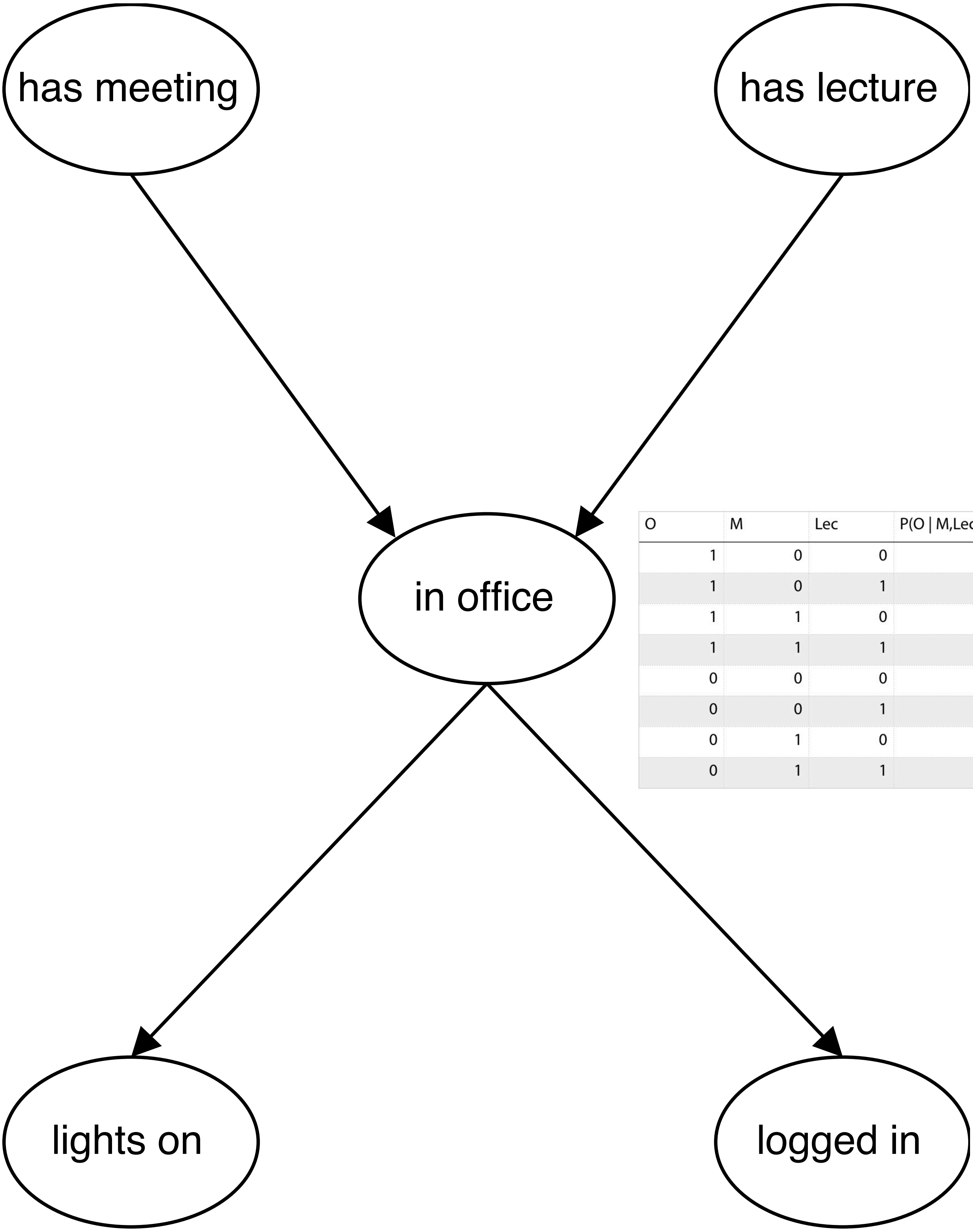
Question One:

Key:

M : has a meeting
Lec : has a lecture
O : Is in the office
Lig : light is on
Log : is logged onto her computer

M	p(M)
0	0.3
1	0.7

L	p(Lec)
0	0.4
1	0.6



O	M	Lec	P(O | M,Lec)
1	0	0	0.06
1	0	1	0.8
1	1	0	0.75
1	1	1	0.95
0	0	0	0.94
0	0	1	0.2
0	1	0	0.25
0	1	1	0.05

Lig	O	P(Lig | O)
0	0	0.98
0	1	0.5
1	0	0.02
1	1	0.5

Log	O	P(Log | O)
0	0	0.8
0	1	0.2
1	0	0.2
1	1	0.8

Question Two:

P(m) * P(lec) * P(o | m, lec) * P(lig | o) * P(log | o)

Using the rule: free parameters of a node = 2^n where n is the number of conditions of the node.

= $2^0 + 2^0 + 2^2 + 2^1 + 2^1$

= 1 + 1 + 4 + 2 + 2

= 10

Question Three:

$P(lec,\overline{m}, o, log,\overline{lig}) = p(lec) * p(\overline{m}) * p(o|lec,\overline{m}) * p(log|o) * p(\overline{lig}|o)$

= 0.6 * 0.3 * 0.8 * 0.8 * 0.5

= 0.0576 (5.76% chance)

Question Four:

$= \sum m \sum lec\ P(O|m, lec)$

= P (O, 0, 0) + P(O, 0, 1) + P (O, 1, 0) + P (O, 1, 1)

= (P(O | 0, 0) * P(m=f) * P(lec=f)) + (P(O | 0, 1) * P(m=f) * P(lec=t)) +

(P(O | 1, 0) * P(m=t) * P(lec=f)) + (P(O | 1, 1) * P(m=t) * P(lec=t))

= (0.06 * 0.3 * 0.4) + (0.8 * 0.3 * 0.6) + (0.75 * 0.7 * 0.4) + (0.95 * 0.7 * 0.6)

= 0.7602 (76.02% chance)

Question Five:

= P(log = t, lig = f | O = t)

= P(log = t | O = t) * P(lig = f) | O = t)

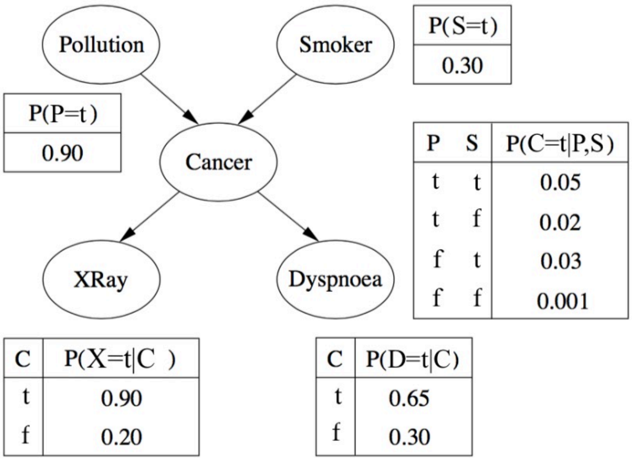
= 0.8 * 0.5

= 0.4 (40% chance)

Question Six:

None, as they are independent of each other. Whether she is logged in or not depends on whether she is in the office or not. Whether her lights are on or not depend on whether she is in the office also. As the are conditionally independent they have no effect on each other.

Part Four: Inference in Bayesian Networks



P(S=t)
0.30

P(P=t)
0.90

P	S	P(C=t|P,S)
t	t	0.05
t	f	0.02
f	t	0.03
f	f	0.001

C	P(X=t|C)
t	0.90
f	0.20

C	P(D=t|C)
t	0.65
f	0.30

Question One:

Part i)

Evidence: P(x)

The evidence is what we already know occurred. In this case we know that the patient had an Xray as X = t.

Hidden: Smoker, Cancer, Dyspnoea

The hidden variables are the ones we are unsure of. In this case we know the evidence is P(X=t) and we are trying to find P(P=t). Therefore that leaves Smoker, Cancer and Dyspnoea as the hidden variables.

Query:

The query is what we are trying to find out. In this case we want to find the probability Pollution is true (P = t).

Part ii)

Step 1: Rewrite in terms of joint distribution

Fix the query and evidence variables (lower case represents constant values, upper case represents variables). Sum over unknown variables and add normalising constant.

$P(p{=}t \mid x{=}t) = P(x, t) / P(t) = \alpha\, P(x,t)$

$= \sum_{S,C,D} P(x,t,S,C,D)$

Step 2: Rewrite joint probability using bayes net factors

Expand the original equation out and re-write it using more specifically. This enables us to use the CPT tables to get the values for each probability.

$= \sum_{S,C,D} P(x,t,S,C,D) = \sum_{S,C,D} P(S)\, P(C \mid p, S)\, P(x \mid C)\, P(D \mid C)$

Step 3: Choose variable order and take summations inside equation

Take the summations inside the equation, which simplifies the equation and enables us to factor variables out later on.

$= P(p) \sum_{S} P(S) \sum_{C} P(C \mid p, S)\, P(x \mid C) \sum_{D} P(D \mid C)$

Step 4: Factor out

$=P(p) \sum_{S} P(S) \sum_{C} P(C \mid p, S)\, P(x \mid C) \sum_{D} P(D \mid C)$

a) Start on the right hand side, and factor variables out working left. First sum out D.

$= P(p) \sum_{S} P(S) \sum_{C} P(C \mid p, S)\, P(x \mid C)\, f_1(C)$

$f_1(C)$

C	P(D | C)
t	0.65
f	0.3

b) Factor out the next variable, C and f1.

$= P(p) \sum_{S} P(S)\ f_2(P,S)$

$f_2(P,S)$

P	S	Calculation	P(C | p, S) P(x | C)
t	t	0.05 * 0.9 + 0.95 * 0.2	0.235
t	f	0.02 * 0.9 + 0.98 * 0.2	0.214
f	t	0.03 * 0.9 + 0.97 * 0.2	0.221
t	t	0.001 * 0.9 + 0.999 * 0.2	0.2007

c) Factor out the final variable, S and f2.

$= \alpha\ P(p)\ f_3(P)$

$f_3(P)$

P	Calculation	P(S)
t	0.9 * 0.235 + 0.1 * 0.214	0.2329
f	0.9 * 0.221 + 0.1 * 0.2007	0.21897

I would eliminate them in order D, S, C as that yields the best result. Ordering them that way means that each summation has the minimum possible values, making variable elimination easier. Please note alpha (α) is calculated in Part iii).

Part iii)

$\alpha = 1 / ((P(p{=}t)\ f_3(P{=}t)) + (P(p{=}f)\ f_3(P{=}f)))$

$\alpha = 1 / ((0.9 * 0.2329) + (0.1 * 0.21897)) = 4.3195238157$

$P(p{=}t \mid x{=}t) = \alpha\ P(p{=}t)\ f_3(P{=}t)$

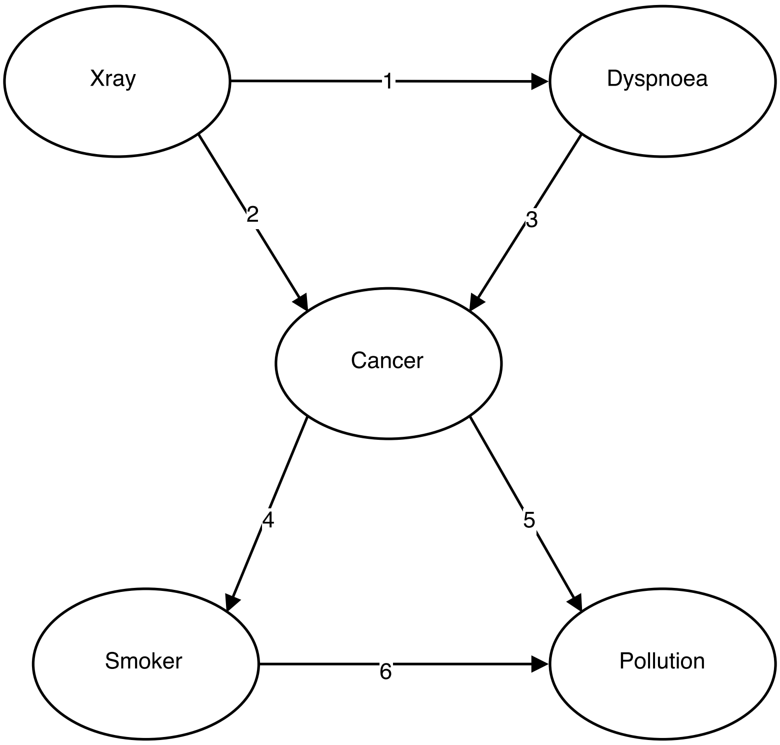
$= \alpha * 0.9 * 0.2329$

$= 4.3195238157 * 0.9 * 0.2329$
$= 0.905415387$

Question Two:

- Pollution and Smoker. They are both independent of each other.
- Xray and Dyspnoea. They are both conditionally independent of each other.
- Xray and Dyspnoea are conditionally independent from Pollution and Smoker given Cancer.

Question Three:



1) We know Xray is conditionally independent from Dyspnoea but we don't know if they are truely independent therefore we need to add the connection.
2) Xray needs a connection to cancer as it is a consequence variable that relates to a cause.
3) Dyspnoea needs a connection to cancer as it is a consequence variable that relates to a cause.
4) Cancer needs a connection to smoker as it is a consequence variable that relates to a cause.
5) Cancer needs a connection to pollution as it is a consequence variable that relates to a cause.
6) We know Smoker is conditionally independent from Pollution but we don't know if they are truely independent therefore we need to add the connection.