COMP307 - Assignment 3:

Uncertainty and Probability

Nicholas Tran, 300296259

Part 1: Reasoning Under Uncertainty Basics

Question 1:

X	$P(X)$
0	0.35
1	0.65

X	Y	$P(Y\|X)$
0	0	0.10
0	1	0.90
1	0	0.60
1	1	0.40

Y	Z	$P(Z\|Y)$
0	0	0.70
0	1	0.30
1	0	0.20
1	1	0.80

1.

Using the **chain rule**: $P(X,\ Y,\ Z) = P(X) \times P(Y|X) \times P(Z \mid X,\ Y)$

And since X and Z are conditionally independent given Y $(X \perp Z \mid Y)$

Then using one of the **conditional independence rules**: $P(Z\ |X,\ Y) = P(Z|Y)$

Rewriting the equation we get: $P(X,\ Y,\ Z) = P(X) \times P(Y|X) \times P(Z|Y)$

$P(X = 0,\ Y = 0,\ Z = 0) = P(X = 0) \times P(Y = 0|X = 0) \times P(Z = 0 \mid Y = 0)$

$= 0.35 \times 0.10 \times 0.70 = 0.0245$

$P(X = 0,\ Y = 0,\ Z = 1) = P(X = 0) \times P(Y = 0|X = 0) \times P(Z = 1 \mid Y = 0)$

$= 0.35 \times 0.10 \times 0.30 = 0.0105$

$P(X = 0,\ Y = 1,\ Z = 0) = P(X = 0) \times P(Y = 1|X = 0) \times P(Z = 0 \mid Y = 1)$

$= 0.35 \times 0.90 \times 0.20 = 0.063$

$P(X = 0,\ Y = 1,\ Z = 1) = P(X = 0) \times P(Y = 1|X = 0) \times P(Z = 1 \mid Y = 1)$

$= 0.35 \times 0.90 \times 0.80 = 0.252$

$P(X=1,\ Y=0,\ Z=0) = P(X=1) \times P(Y=0|X=1) \times P(Z=0 \mid Y=0)$

$= 0.65 \times 0.60 \times 0.70 = 0.273$

$P(X=1,\ Y=0,\ Z=1) = P(X=1) \times P(Y=0|X=1) \times P(Z=1 \mid Y=0)$

$= 0.65 \times 0.60 \times 0.30 = 0.117$

$P(X=1,\ Y=1,\ Z=0) = P(X=1) \times P(Y=1|X=1) \times P(Z=0 \mid Y=1)$

$= 0.65 \times 0.40 \times 0.20 = 0.052$

$P(X=1,\ Y=1,\ Z=1) = P(X=1) \times P(Y=1|X=1) \times P(Z=1 \mid Y=1)$

$= 0.65 \times 0.40 \times 0.80 = 0.208$

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.0245
0	0	1	0.0105
0	1	0	0.063
0	1	1	0.252
1	0	0	0.273
1	0	1	0.117
1	1	0	0.052
1	1	1	0.208

2.

Using the **product rule**: $P(X,\ Y) = P(X) \times P(Y|X)$

$P(X=0,\ Y=0) = P(X=0) \times P(Y=0|X=0)$

$= 0.35 \times 0.10 = 0.035$

$P(X=0,\ Y=1) = P(X=0) \times P(Y=1|X=0)$

$= 0.35 \times 0.90 = 0.315$

$P(X=1,\ Y=0) = P(X=1) \times P(Y=0|X=1)$

$= 0.65 \times 0.60 = 0.39$

$P(X=1,\ Y=1) = P(X=1) \times P(Y=1|X=1)$

$= 0.65 \times 0.40 = 0.26$

X	Y	$P(X,Y)$
0	0	0.035
0	1	0.315
1	0	0.39
1	1	0.26

3.

Using the **sum rule** where the marginal probability is the sum over all it's joint probabilities.

(a) $P(Z=0)$

$P(Z=0) = P(X=0,\ Y=0,\ Z=0) + P(X=0,\ Y=1,\ Z=0) + P(X=1,\ Y=0,\ Z=0) + P(X=1,\ Y=1,\ Z=0)$

$= 0.0245 + 0.063 + 0.273 + 0.052$

$= 0.4125$

(b) $P(X = 0,\ Z = 0)$

$P(X = 0,\ Z = 0) = P(X = 0,\ Y = 0,\ Z = 0) + P(X = 0,\ Y = 1,\ Z = 0)$

$= 0.0245 + 0.063$

$= 0.0875$

(c) $P(X = 1, Y = 0|Z = 1)$

$P(X = 1, Y = 0|Z = 1) = P(X = 1,\ Y = 0,\ Z = 1) = 0.117$

(d) $P(X = 0|Y = 0, Z = 0)$

$P(X = 0|Y = 0, Z = 0) = P(X = 0,\ Y = 0,\ Z = 0) = 0.0245$

Question 2:

- $P(B = t) = 0.7$
- $P(C = t) = 0.4$
- $P(A = t|B = t) = 0.3$
- $P(A = t|C = t) = 0.5$
- $P(B = t|C = t) = 0.2$

(i) $P(B = t, C = t)$

$P(B = t, C = t) = P(B = t) \times P(C = t|B = t) = P(C = t) \times P(B = t|C = t) = 0.4 * 0.2 = 0.08$

(ii) $P(A = f, B = t)$

$P(A = f, B = t) = P(A = f) \times P(B = t|A = f) = P(B = t) \times P(A = f|B = t)$

$P(A = f|B = t) = 1 - P(A = t|B = t) = 1 - 0.3 = 0.7$

$P(B = t) \times P(A = f|B = t) = 0.7 * 0.7 = 0.49$

(iii) $P(A = t, B = t|C = t)$

$P(A = t, B = t|C = t) = P(A = t|C = t) \times P(B = t|C = t) = 0.5 * 0.2 = 0.1$

(iv) $P(A = t|B = t, C = t)$

$P(A = t|B = t, C = t) = P(A = t|C = t) = 0.5$

(v) $P(A = t, B = t, C = t)$

$P(A = t, B = t, C = t) = P(C = t) \times P(B = t|C = t) \times P(A = t|B = t, C = t) = 0.4 * 0.2 * 0.5 = 0.04$

Part 2: Naive Bayes Method

	P(Y=no-recurrence-events)	P(Y=recurrence-events)
P(Y=y)	71.16%	29.59%
P(age = 30-39|Y=y)	11.58%	20.25%
P(age = 40-49|Y=y)	32.63%	34.18%
P(age = 60-69|Y=y)	20.00%	21.52%
P(age = 50-59|Y=y)	34.21%	27.85%
P(age = 70-79|Y=y)	3.16%	1.27%
P(age = 20-29|Y=y)	1.05%	1.27%
P(menopause = premeno|Y=y)	51.58%	62.03%
P(menopause = ge40|Y=y)	46.32%	39.24%
P(menopause = lt40|Y=y)	3.16%	1.27%
P(tumor-size = 30-34|Y=y)	17.89%	29.11%
P(tumor-size = 20-24|Y=y)	18.42%	17.72%
P(tumor-size = 15-19|Y=y)	12.11%	8.86%
P(tumor-size = 0-4|Y=y)	4.21%	2.53%
P(tumor-size = 25-29|Y=y)	16.84%	24.05%
P(tumor-size = 50-54|Y=y)	2.63%	5.06%
P(tumor-size = 10-14|Y=y)	13.68%	2.53%
P(tumor-size = 40-44|Y=y)	8.95%	8.86%
P(tumor-size = 35-39|Y=y)	6.32%	10.13%
P(tumor-size = 5-9|Y=y)	2.63%	1.27%
P(tumor-size = 45-49|Y=y)	1.58%	2.53%
P(inv-nodes = 0-2|Y=y)	84.74%	54.43%
P(inv-nodes = 6-8|Y=y)	4.21%	13.92%
P(inv-nodes = 9-11|Y=y)	1.58%	7.59%
P(inv-nodes = 3-5|Y=y)	8.95%	20.25%
P(inv-nodes = 15-17|Y=y)	2.11%	5.06%
P(inv-nodes = 12-14|Y=y)	1.05%	3.80%
P(inv-nodes = 24-26|Y=y)	0.53%	2.53%
P(node-caps = no|Y=y)	87.89%	60.76%

Nicholas Tran, 300296259

P(node-caps = yes|Y=y)	12.63%	40.51%
P(deg-malig = 3|Y=y)	20.00%	54.43%
P(deg-malig = 2|Y=y)	51.58%	36.71%
P(deg-malig = 1|Y=y)	29.47%	11.39%
P(breast = left|Y=y)	51.05%	55.70%
P(breast = right|Y=y)	49.47%	45.57%
P(breast-quad = left_low|Y=y)	37.37%	40.51%
P(breast-quad = right_up|Y=y)	11.05%	17.72%
P(breast-quad = left_up|Y=y)	35.26%	31.65%
P(breast-quad = right_low|Y=y)	9.47%	8.86%
P(breast-quad = central|Y=y)	8.95%	6.33%
P(irradiat = no|Y=y)	84.74%	62.03%
P(irradiat = yes|Y=y)	15.79%	39.24%

	class	**prediction**	**no-recurrence-events probability**	**recurrence-events probability**
189	no-recurrence-events	recurrence-events	0.0012%	0.0005%
190	no-recurrence-events	no-recurrence-events	0.0417%	0.0043%
191	no-recurrence-events	no-recurrence-events	0.0059%	0.0002%
192	no-recurrence-events	no-recurrence-events	0.0190%	0.0017%
193	no-recurrence-events	no-recurrence-events	0.0005%	0.0003%
194	no-recurrence-events	no-recurrence-events	0.0750%	0.0065%
195	no-recurrence-events	no-recurrence-events	0.0259%	0.0124%
274	recurrence-events	no-recurrence-events	0.0394%	0.0015%
275	recurrence-events	recurrence-events	0.0108%	0.0049%
276	recurrence-events	recurrence-events	0.0088%	0.0051%

Test Accuracy: 80.0%

Nicholas Tran, 300296259

Part 3: Building Bayesian Network

1.

<u>Random Variables:</u>
Meeting: Whether or not Rachel has a research meeting with her postgraduate students.
Lectures: Whether or not Rachel has lectures to teach her undergraduate students.
Office: Whether Rachel comes into the office or works from home.
Light On: Whether or not Rachel has her office light on.
Computer On: Whether or not Rachel is logged onto her work computer.

<u>Causal Dependencies:</u>
- Rachel prefers to work from home so she only comes into the office if she has a meeting or lectures to teach or both
- The light in her office is only turn on half of the time to hide from others when she needs to get work done.
- Rachel is logged into her work computer when she is in the office and sometimes when she is working from home.

<u>Conditional Probabilities:</u>
- The probability that Rachel has a meeting is 70%
- The probability that Rachel has lectures to teach is 60%
- The probability that Rachel comes into the office if she has a meeting but no lectures is 75%
- The probability that Rachel comes into the office if she has lectures but not meetings is 80%
- If Rachel has neither meetings or lectures then the probability she comes into the office is 6%
- If Rachel is in her office then she has the light on half of the time (50%)
- If Rachel is not in her office then she only has the light on 2% of the time for cleaners
- If Rachel is in the office then 80% of the time she is logged into her computer
- If Rachel is not in her office and since she often works from home then the probability she's logged into her computer is 20%

Given the node order of: Meeting, Lectures, Office, Light On, Computer On

Step 1: Add node Meeting
Step 2: Add node Lectures
- P(Lectures|Meeting) = P(Lectures)? Yes, no link

Step 3: Add node Office
- P(Office|Meeting,Lectures) = P(Office)? No

- P(Office|Meeting,Lectures) = P(Office|Lectures)? No
- P(Office|Meeting,Lectures) = P(Office|Meeting)? No, Meeting -> Office and Lectures -> Office

Step 4: Add node Light On

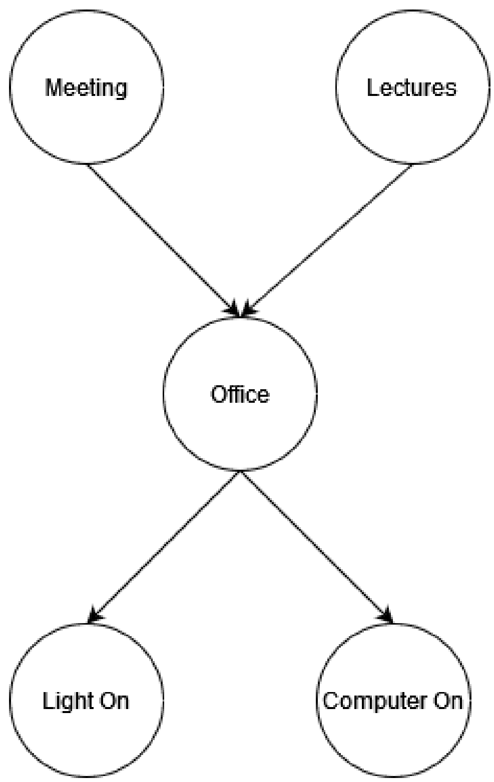
- P(Light On|Meeting,Lectures, Office) = P(Light On|Office)? Yes, Office -> Light On, no link from Meeting or Lectures to Light On.

Step 5: Add node Computer On

- P(Computer On|Meeting,Lectures,Office) = P(Computer On|Office)? Yes, Office -> Computer On, no link from Meeting or Lectures to Computer On

P(Meeting=t)
0.7

P(Lectures=t)
0.6



Meeting	Lectures	P(Office|Meeting,Lectures)
t	t	0.95
t	f	0.75
f	t	0.8
f	f	0.06

Office	P(Light On|Office)
t	0.5
f	0.02

Office	P(Computer On|Office)
t	0.8
f	0.2

Nicholas Tran, 300296259

2.

Number of Free Parameters:
|Meeting| = 2, |Lectures| = 2, |Office| = 2, |Light On| = 2, |Computer On| = 2
Meeting has no parent so the number of free parameters is 2-1=1. Same for Lectures.
Office has two parents leading to 4 possible conditions, so it's free parameters is 1*2*2=4.
Light On and Computer On have a common cause so their free parameters are 2 and 2.
So in total the number of free parameters is 1+1+4+2+2=10

3.

P(Lectures=t, Meeting=f, Office=t, Computer On=t, Light On=f) =

P(Lectures=t)*P(Meeting=f)*P(Office|Lectures=t,Meeting=f)*P(Computer On=t|Office=t)*P(Light On=f|Office=t) =

0.6*0.3*0.8*0.8*0.5=0.0576

4.

P(Office=t) =

P(Office=t, Meeting=t, Lectures=t)
+P(Office=t, Meeting=t, Lectures=f)
+P(Office=t, Meeting=f, Lectures=t)
+P(Office=t, Meeting=f, Lectures=f) =

P(Office=t|Meeting=t,Lectures=t)*P(Meeting=t)*P(Lectures=t)
+P(Office=t|Meeting=t,Lectures=f)*P(Meeting=t)*P(Lectures=f)
+P(Office=t|Meeting=f,Lectures=t)*P(Meeting=f)*P(Lectures=t)
+P(Office=t|Meeting=f,Lectures=f)*P(Meeting=f)*P(Lectures=f) =

0.95*0.7*0.6
+0.75*0.7*0.4
+0.8*0.3*0.6
+0.06*0.3*0.4 = 0.7602 (76.02%)

5.

P(Computer On=t,Light On=f|Office=t) = P(Computer On=t|Office=t)*P(Light On=f|Office=t)=0.8*0.5=0.4

Part 4: Inference in Bayesian Networks

1.

Evidence: X-Ray

Hidden: Smoker, Cancer, Dyspnoea

Query: Pollution

2.

Product rule:

$P(P|x) = \frac{P(B,x)}{P(x)}$

$\alpha = \frac{1}{P(x)}$

Sum rule:

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

Factorisation:

$= \sum_{S,C,D} P(P)P(S)P(C|P,S)P(X|C)P(D|C)$

Define initial factors:

$f_1(P) = P(P)$:

P	$P(P)$
t	0.90
f	0.10

$f_2(S) = P(S)$:

S	$P(S)$
t	0.30
f	0.70

$f_3(C,P,S) = P(C|P,S)$:

C	P	S	$P(C\|P,S)$
t	t	t	0.05
f	t	t	0.95
t	t	f	0.02
f	t	f	0.98
t	f	t	0.03
f	f	t	0.97
t	f	f	0.001
f	f	f	0.999

$f_4(X, C) = P(x|C)$:

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

$f_5(D, C) = P(D|C)$:

D	C	$P(D|C)$
t	t	0.65
t	f	0.30
f	t	0.35
f	f	0.70

Eliminate hidden: Dysponea -> Cancer -> Smoker

Eliminate Dysponea:

D	C	$P(D|C)$
t	t	0.65
t	f	0.30
f	t	0.35
f	f	0.70

C	...
t	1
f	1

Eliminate Cancer by joining $f_4(X, C) = P(x|C)$ and $f_3(C, P, S) = P(C|P, S)$:

C	P	S	$P(C|P, S)$
t	t	t	0.05
f	t	t	0.95
t	t	f	0.02
f	t	f	0.98
t	f	t	0.03
f	f	t	0.97
t	f	f	0.001
f	f	f	0.999

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

C	P	S	...
t	t	t	0.05*0.90
f	t	t	0.95*0.20
t	t	f	0.02*0.90
f	t	f	0.98*0.20
t	f	t	0.03*0.90
f	f	t	0.97*0.20
t	f	f	0.001*0.90
f	f	f	0.999*0.20

P	S	...
t	t	(0.05*0.90)+(0.95*0.20)=0.235
t	f	(0.02*0.90)+(0.98*0.20)=0.214
f	t	(0.03*0.90)+(0.97*0.20)=0.221
f	f	(0.001*0.90)+(0.999*0.20)=0.200

Eliminate Smoker:

P	S	...
t	t	0.235
t	f	0.214
f	t	0.221
f	f	0.200

S	$P(S)$
t	0.30
f	0.70

P	S	...
t	t	0.235*0.30
t	f	0.214*0.70
f	t	0.221*0.30
f	f	0.200*0.70

P	...
t	(0.235*0.30)+(0.214*0.70)
f	(0.221*0.30)+(0.200*0.70)

P	...
t	0.2203
f	0.2063

Join previous table with $f_1(P) = P(P)$:

P	...
t	0.2203
f	0.2063

P	$P(P)$
t	0.90
f	0.10

P	...
t	0.2203*0.90
f	0.2063*0.10

P	$P(P,x)$
t	0.19827
f	0.02063

Final table:

P	$P(P\|x)$
t	$\frac{0.19827}{0.19827+0.02063}$
f	$\frac{0.02063}{0.19827+0.02063}$

P	$P(P\|x)$
t	0.90575605299
f	0.094243947

3.

Final Probability = 0.90575605299 or 90.58%