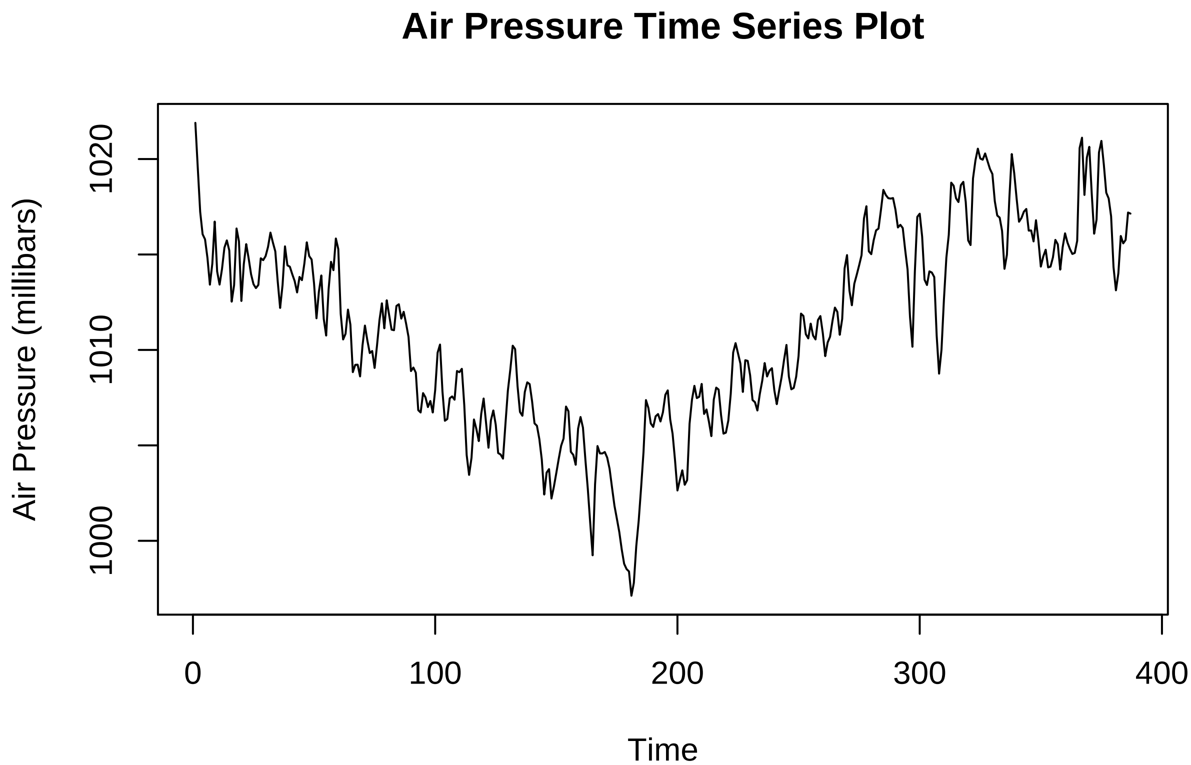
SIT743 Assignment 2

Student Name: Dhananjay Pandya | Student Number: 218202943

Question 1)

1.1)

Time Series Plot

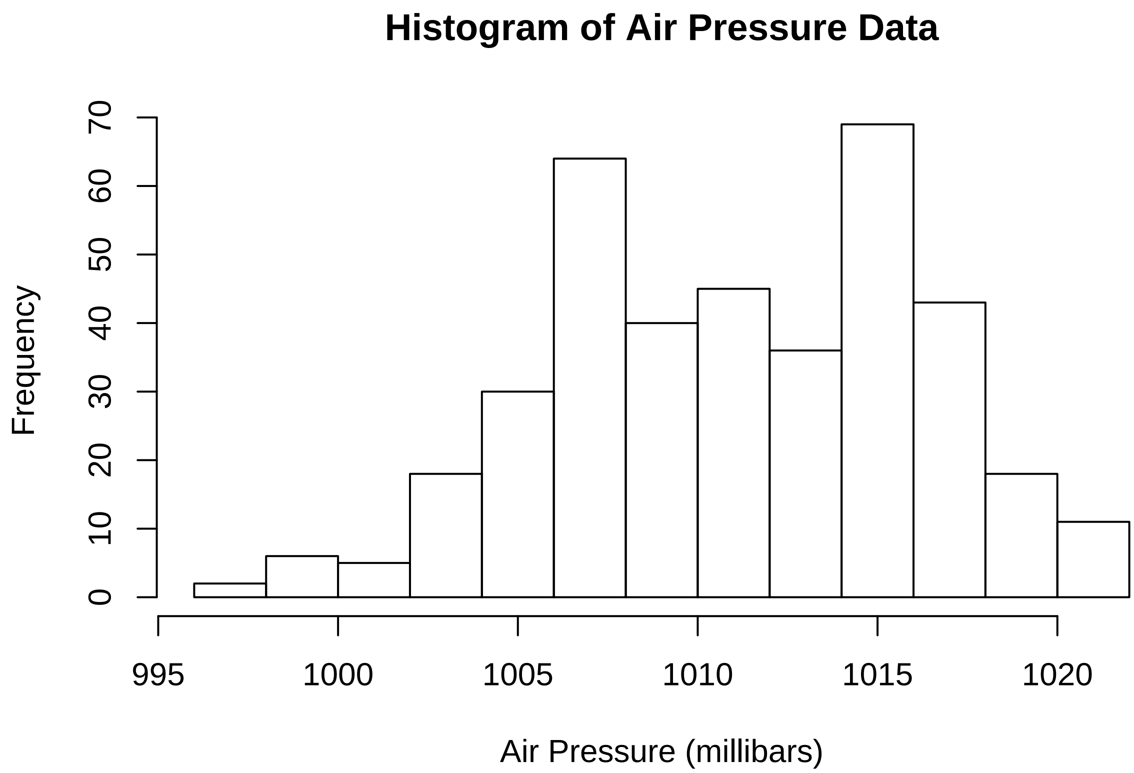


Five Point Summary

Measure	Value
Minimum	997.1
1st Quartile	1007.2
Median	1011.4
3rd Quartile	1015.3
Maximum	1021.9

Mean: 1011.14
Standard Deviation: 5.147

1.2)



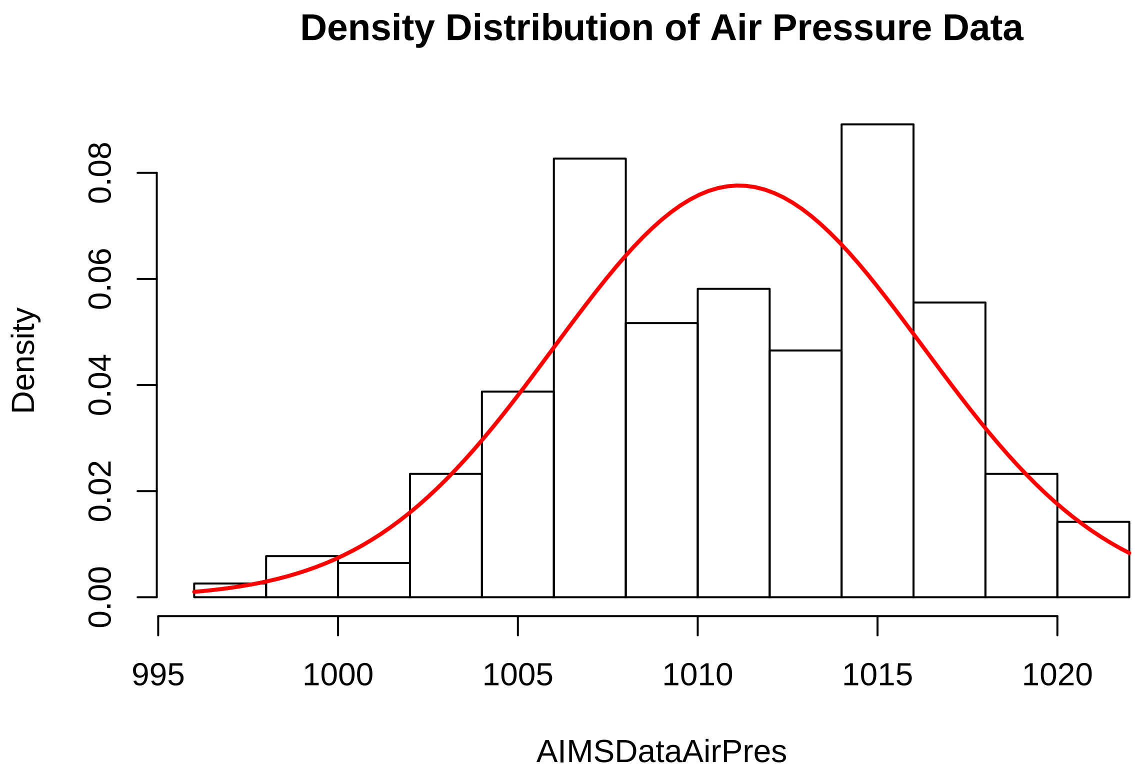
Comments:

1. Shape: The shape of this histogram is bimodal. This means that there are two peaks occuring at ~1007 millibars and ~1015 millibars. This is because air pressure usually varies between high pressure and low pressure, which results in localised weather changes.

2. Span: This dataset spans ~25 millibars. This is to be expected because air pressure is constrained on earth.

3. Outliers: There are some very low air pressure measurements. These can be seen on the left hand side of the histogram. This could be because the region experienced extreme low weather situations such as storms.

1.3)

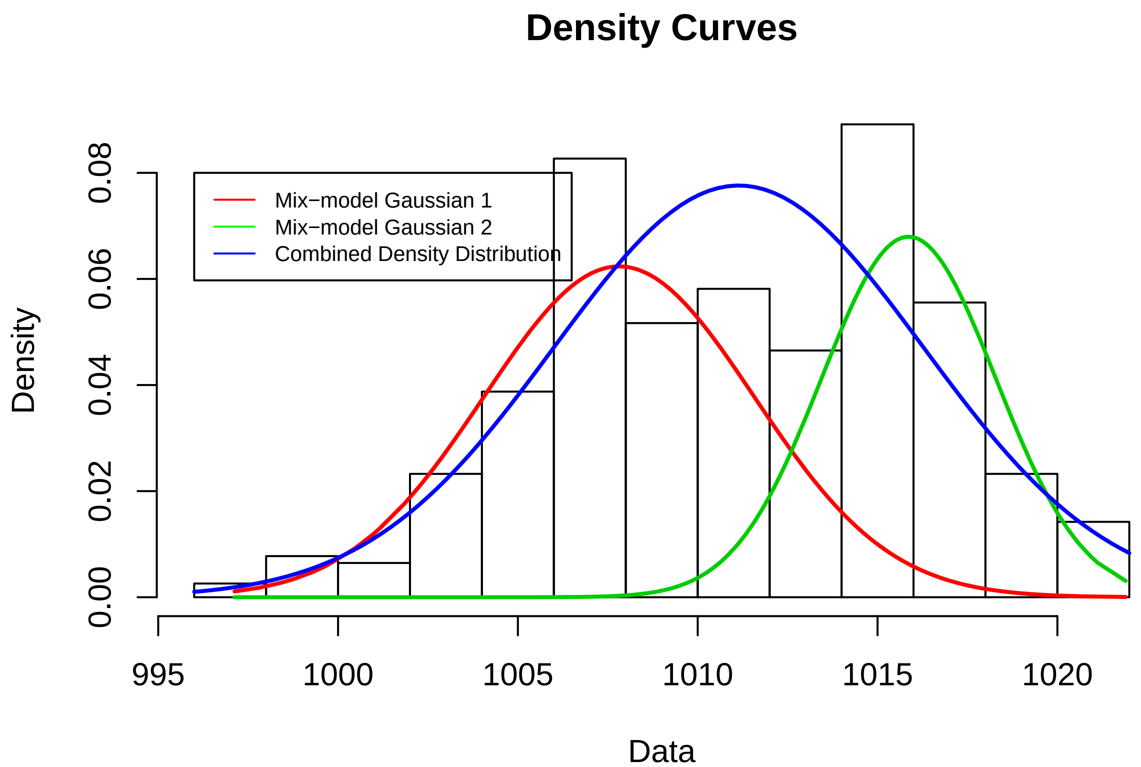


Mean: 1011.14
Standard Deviation: 5.140

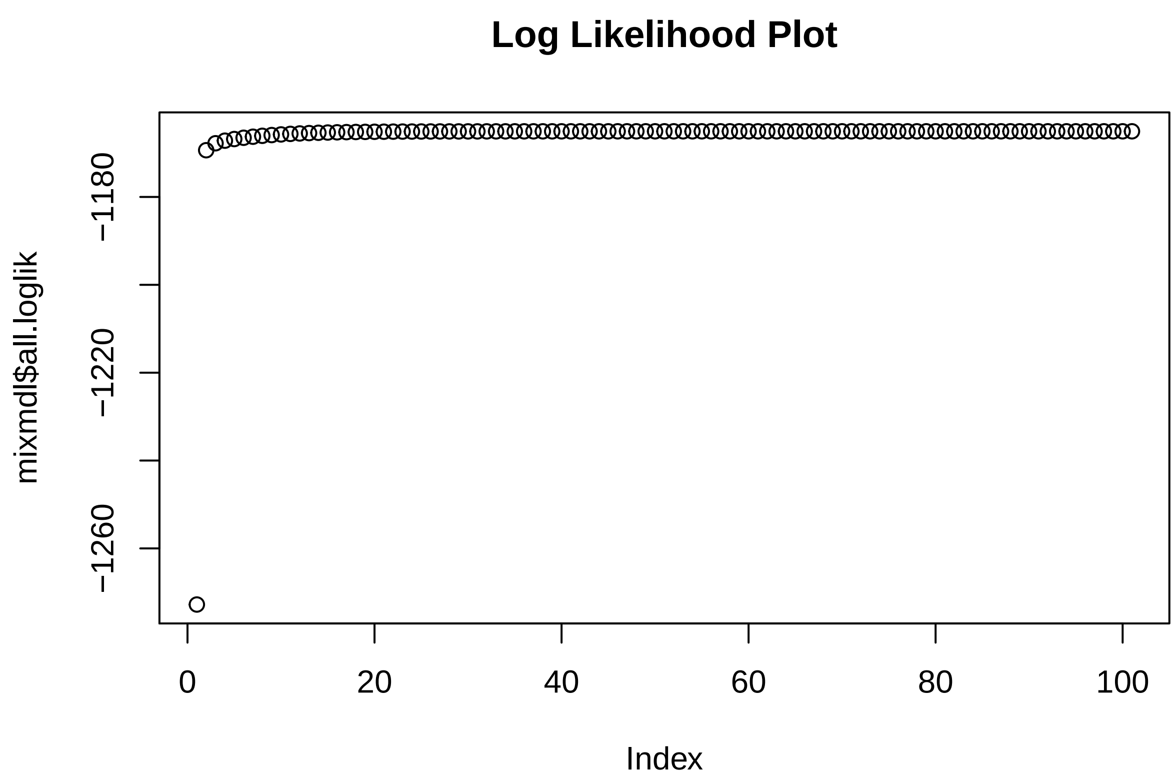
1.4)

Measure	Gaussian 1	Gaussian 2
Mixing Coefficient	0.586942	0.413058
Mean	1007.813192	1015.864975
Standard Deviation	3.754725	2.425176

1.5)


Density Curves
Mix−model Gaussian 1
Mix−model Gaussian 2
Combined Density Distribution
Density
0.00
0.02
0.04
0.06
0.08
995
1000
1005
1010
1015
1020
Data

1.6)


Log Likelihood Plot
mixmdl$all.loglik
−1260
−1220
−1180
0
20
40
60
80
100
Index

1.7)

Comment: The mixture of Gaussians is better than the single Gaussian because it is a better representation of air pressure clustering. As mentioned previously, air pressures vary between low and high values. The single Gaussian places the center of the distribution between the two modes that are shown on the histogram. The mixture of Gaussians is more precise in that the center of distribution 1 and 2 coincide with the modes of the histogram.

1.8)

Problem: Severe overfitting resulting from the presence of singularities

Resolution: Use experience to detect this issue. Reset the mean to a randomly chosen value while resetting covariance to some large value and then continue with the optimisation.

Question 2)

2.1)

$$p(A,B,C,D,E) = p(A)p(B)p(C|A,B)p(D|C)p(E|C)$$

2.2)

Parameters to fully specify A: 3
Parameters to fully specify B: 1
Parameters to fully specify C: 18
Parameters to fully specify D: 6
Parameters to fully specify E: 4
$Total = 3 + 1 + 18 + 6 + 4 = 31$

2.3)

Parameters required if no independencies among variables: $= (4 \times 2 \times 3 \times 3 \times 2) - 1 = 143$

Comments: If independence cannot be assumed then the number of parameters required grows exponentially. Bayesian networks are useful because they create independencies between variable. As a result the number of parameters grows linearly. This is evident because in the non-independence case, the number of parameters required to fully specify the distribution is 143, whereas for the given Bayesian network, the number of parameters required to fully specify the distribution is 31.

2.4)

$$p(A = summer|D = dark, E = wide) = \frac{p(A{=}summer, D{=}dark, E{=}wide)}{p(D{=}dark, E{=}wide)} \tag{1}$$

$$= \frac{\sum_{B,C} p(A{=}summer)p(B)p(C|A{=}summer,B)p(D{=}dark|C)p(E{=}wide|C)}{\sum_{A,B,C} p(A)p(B)p(C|A,B)p(D{=}dark|C)p(E{=}wide|C)} \tag{2}$$

From this equation, we can elimate $p(B)$, $p(D = dark|C)$, and $p(E = wide|C)$ because they appear in the numerator and the denominator. This leaves us with the equation to compute the required conditional probability.

$$p(A = summer|D = dark, E = wide) = \frac{\sum_{B,C} p(A = summer)p(C|A = summer, B)}{\sum_{A,B,C} p(A)p(C|A,B)} \tag{3}$$

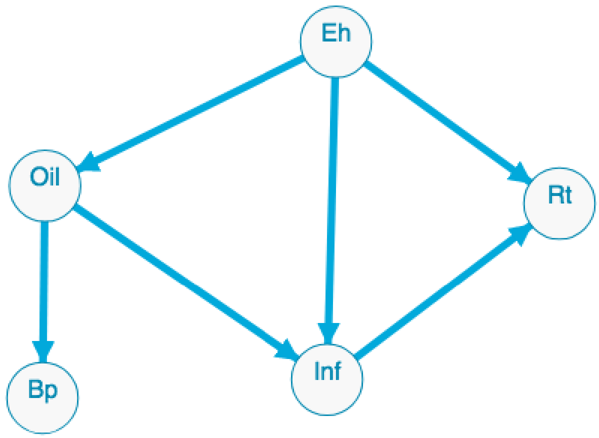


Figure 1: Belief network for given distribution

Question 3)

3.1)

See Figure 1

3.2)

```
## $eh
## eh
##  low high
##  0.2  0.8
## attr(,"class")
## [1] "parray" "array"
##
## $oil
##       eh
## oil    low high
##   low  0.9 0.05
##   high 0.1 0.95
## attr(,"class")
## [1] "parray" "array"
##
```

```
## $inf
## , , oil = low
##
##       eh
## inf    low high
##   low  0.9  0.2
##   high 0.1  0.8
##
## , , oil = high
##
##       eh
## inf    low high
##   low  0.1 0.02
##   high 0.9 0.98
##
## attr(,"class")
## [1] "parray" "array"
##
## $bp
##         oil
## bp        low high
##   low    0.80  0.1
##   high   0.15  0.4
##   normal 0.05  0.5
## attr(,"class")
## [1] "parray" "array"
##
## $rt
## , , inf = low
##
##         eh
## rt       low high
##   low    0.6  0.1
##   high   0.3  0.2
##   normal 0.1  0.7
##
## , , inf = high
##
##         eh
## rt       low high
##   low    0.2 0.05
##   high   0.2 0.10
##   normal 0.6 0.85
##
## attr(,"class")
## [1] "parray" "array"
```

3.3)

$p(inf = high|bp = high, rt = normal) = 0.9719$

Therefore, the probability is ~97.19% as computed by R.

Question 4)

4.1)

Overall concept of d-separation: Two nodes A and B are d-separated given C, if all paths from A to B given the set C are blocked. If this is the case then A is conditonally independent of B given the set C.

a)

C is not guaranteed to be marginally independent of G. This is because G is a direct descendent of G and thus, knowing G does give us some information about C.

b)

Given E, H is not guaranteed to be conditionally independent of C. This is because all paths from C to H are not blocked. The active path from C to H passes through node F (head-to-tail node and F is not given).

c)

Given D, G is guaranteed to be conditionally independent of E. All paths from G to E encounter node F (head-to-head node with neither F nor H in the given set) and are blocked.

d)

Given F, H is not guaranteed to be conditionally independent of C. This is because all paths from C to H are not blocked. One active path from C to H passes through nodes A (tail-to-tail and A is not given), D (head-to-tail and D is not given), F (head-to-head w.r.t to node E, and F is given) and E (tail-to-tail node and E is not given).

e)

Given F, G is not guaranteed to be conditionally independent of B. This is because all paths from B to H are not blocked. One active path from B to H passes through nodes D (head-to-head node with descendent F given), F (head-to-head node w.r.t to node E, and F is given) and E (tail-to-tail node and E is not given).

f)

Given the set {D,C,E}, B is guaranteed to be conditionally independent of G. This is because all paths from B to G are blocked based on the given set. The reasoning for each path is as follows:

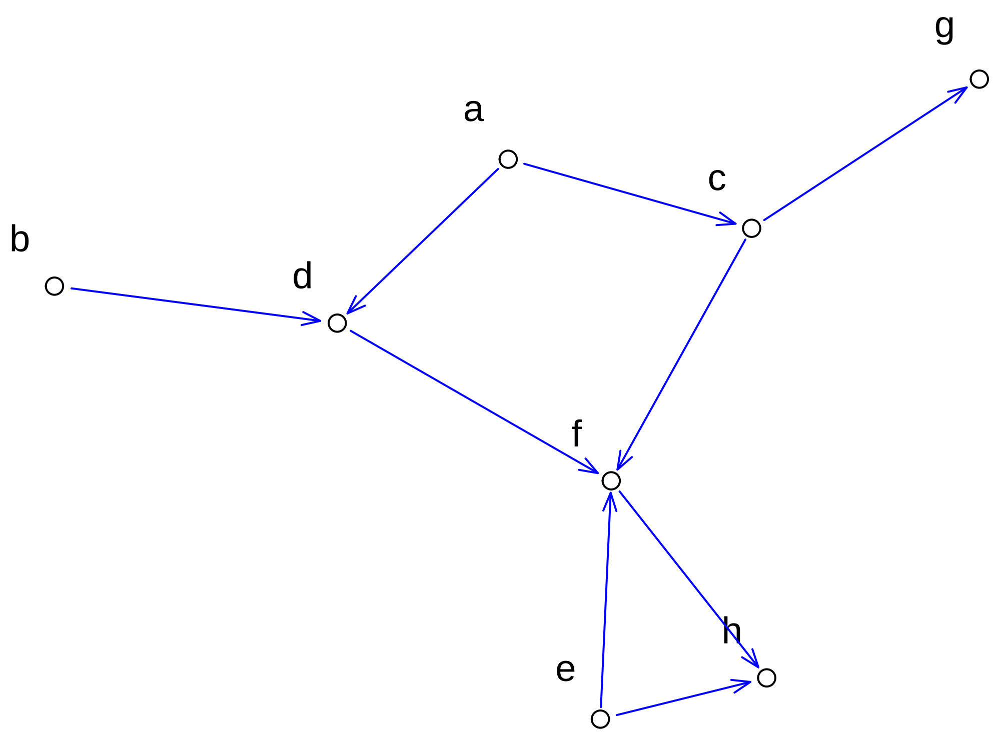
- Although the head-to-head node at D is not blocked, since D is given, the head-to-head node at F is blocked because neither F nor H are given. As a result, the path from B to G, given D is blocked.
- The head-to-head node at D is not blocked, since D is given, and the tail-to-tail node at A is not blocked because A is not given. However, the head-to-tail node at C is blocked because C is given. As a result, the path from B to G, given C is blocked.
- The path from B to G given E is also blocked because it encounters node F along the way.

g)

Given the set {D,F}, A is not guaranteed to be conditionally independent of H. This is because all paths from A to H are not blocked. One active path from A to H passes through node D (head-to-head and D and F are given) and node F (head-to-head and F is given).

4.2)

```r
library(igraph)
library(ggm)
dag <- DAG(g~c,c~a,f~c,f~d,d~a,d~b,f~e,h~f,h~e)
drawGraph(dag,adjust=FALSE)
```




```r
dSep(dag,first="c",second="g",cond=NULL)
```

```
## [1] FALSE
```

```r
dSep(dag,first="c",second="h",cond=c("e"))
```

```
## [1] FALSE
```

```r
dSep(dag,first="g",second="e",cond=c("d"))
```

```
## [1] TRUE
```

```r
dSep(dag,first="c",second="h",cond=c("f"))
```

```
## [1] FALSE
```

```r
dSep(dag,first="b",second="g",cond=c("f"))
```

```
## [1] FALSE
```

```r
dSep(dag,first="b",second="g",cond=c("d","c","e"))
```

```
## [1] TRUE
```

```r
dSep(dag,first="a",second="h",cond=c("d","f"))
```

```
## [1] FALSE
```

Question 5)

5.1)

$$
\begin{aligned}
p(D=1|A=0) &= \frac{p(D=1,A=0)}{p(A=0)} && (4)\\
&= \frac{\sum_{B,C} p(D=1,B,C,A=0)}{\sum_{B,C,D} p(D,B,C,A=0)} && (5)\\
&= \frac{\sum_{B,C} p(A=0)p(B|A=0)p(C|A=0,B)p(D=1|C)}{\sum_{B,C,D} p(A=0)p(B|A=0)p(C|A=0,B)p(D|C)} && (6)
\end{aligned}
$$

From this, we can eliminate $p(A=0)$, $p(B|A=0)$, and $p(C|A=0,B)$ because these appear in both the numerator and denominator.

$$
\begin{aligned}
p(D=1|A=0) &= \frac{\sum_{C} p(D=1|C)}{\sum_{C,D} p(D|C)} && (7)\\
&= \frac{(1-\gamma)+0.7}{\gamma+(1-\gamma)+0.3+0.7} && (8)\\
&= \frac{1.7-\gamma}{2} && (9)
\end{aligned}
$$

As shown above, $p(D=1|A=0)$ depends only on γ.

5.2)

$$
\begin{aligned}
p(D=1|A=0) &= \frac{1.7-\gamma}{2} && (10)\\
&= \frac{1.7-0.1}{2} && (11)\\
&= 0.8 && (12)
\end{aligned}
$$

Question 6)

6.1)

$$\begin{aligned} p(F|A=1) &= \frac{p(F, A=1)}{p(A=1)} && (13) \\ & && (14) \end{aligned}$$

Where,

$$\begin{aligned} p(F, A) &= \sum_{B,C,D,E} p(A, B, C, D, E, F) && (15) \\ &= \sum_{B,C,D,E} p(A)p(B|A)p(D)p(C|A, D)p(E|B, C)p(F|E, D) && (16) \\ &= \sum_{B,C,D,E} f_0(A)f_1(A, B)f_2(D)f_3(A, C, D)f_4(B, C, E)f_5(D, E, F) && (17) \\ & && (18) \end{aligned}$$

Given that $A = 1$,

$$\begin{aligned} p(F, A=1) &= \sum_{B,C,D,E} f_1(B)f_2(D)f_3(C, D)f_4(B, C, E)f_5(D, E, F) && (19) \\ & && (20) \end{aligned}$$

6.2)

Eliminate B,

$$\begin{aligned} p(F, A=1) &= (\sum_{C,D,E} f_2(D)f_3(C, D)f_5(D, E, F))(\sum_{B} f_1(B)f_4(B, C, E)) && (21) \\ &= \sum_{C,D,E} f_2(D)f_3(C, D)f_5(D, E, F)f_6(C, E) && (22) \end{aligned}$$

Eliminate C,

$$\begin{aligned} p(F, A=1) &= (\sum_{D,E} f_2(D)f_5(D, E, F))(\sum_{C} f_3(C, D)f_6(C, E)) && (23) \\ &= \sum_{C,D,E} f_2(D)f_5(D, E, F)f_7(D, E) && (24) \end{aligned}$$

Eliminate D,

$$\begin{aligned} p(F, A=1) &= \sum_{D,E} f_2(D)f_5(D, E, F)f_7(D, E) && (25) \\ &= \sum_{E} f_8(E, F) && (26) \end{aligned}$$

Eliminate E,

$$
\begin{aligned}
p(F, A=1) &= \sum_E f_8(E, F) && (27) \\
&= f_9(F) && (28)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
p(F|A=1) &= \frac{p(F, A=1)}{p(A=1)} && (29) \\
&= \frac{f_9(F)}{\sum_F f_9(F)} && (30)
\end{aligned}
$$

Question 7)

Real World Application 1

Bayesian classification has an interesting application in weather prediction. Future rainfall can be predicted using features such as temperature, pressure, humidity and windspeed, along with known quantities of rainfall (labels). This is a form of supervised machine learning where prior probabilities are computed for each class of training data. For all features, the conditional probabilities ($p(Feature|Class)$ are computed. This is used to classify unseen data.

Reference: Janbandhu, CC., Meshram, PD. & Gedam, MN. 2017, 'Modelling Rainfall Prediction Using Data Mining Method - A Bayesian Approach', *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol. 3, no. 11, retrieved 22 September 2018, https://doi.org/10.1109/CIMSim.2013.29

Real World Application 2

Bayesian networks are also used to make inferences from sensor data. One interesting subset of this is the detection of Mastitis in cows. This disease affects the capacity of cows to produce milk. Using data from sensors, a Bayesian network is used to find the probability that a cow or herd has the disease given information from sensors. This helps early detection of the disease and helps farmers avoid revenue loss from milk production along with reducing medical costs.

Reference: Drury, B., Valverde-Rebaza, J., Moura, MF. & Lopes, AA. 2017, 'A survey of the applications of Bayesian networks in agriculture', *Engineering Applications of Artificial Intelligence*, vol. 65, pp.29-42, retrieved 22 September 2018, https://doi.org/10.1016/j.engappai.2017.07.003