STAT650 F2021 HW3 Solution

Xiaomeng Yan

9/21/2021

Problem 1

Airports have installed detection systems that determine whether a bag contains explosives or not. The system has a 95% chance of correctly indentifying a bag containing explosives, and a 99.5% chance of correctly classifying a bag without explosives as safe. A particular airport screens 3 million bags per year and it anticipated that of these bags, 10 will contain explosives.

a. What is the change that a bag identified as containing explosives actually does contain explosives?

Solution:

Define the events as follows:

- $+$: "Postivie Detection".
- $-$: "Negative Detection".
- E: "Contains Explosives".
- nE: "Does not contain explosives".

Then we are given: $P(+|E) = 0.95$, $P(-|nE) = 0.995$, $P(E) = \frac{10}{3{,}000{,}000} = \frac{1}{300{,}000}$.

The problem translates to $P(E|+)$. By Baye's Theorem, this equals:

$$\begin{aligned} P(E|+) &= \frac{P(E)P(+|E)}{P(+)} \\ &= \frac{P(E)P(+|E)}{P(+)} \\ &= \frac{P(E)P(+|E)}{P(+|E)P(E) + P(+|nE)P(nE)} \\ &= \frac{\frac{1}{300{,}000}\cdot 0.95}{\frac{1}{300{,}000}\cdot 0.95 + \frac{299{,}999}{300{,}000}(1-0.995)} \approx 0.00063 \end{aligned}$$

b. If we wanted this probability to be at least 0.5. then what should the probability of correctly identifying a bag not containing explosives need to be?

Solution:

We want this probability to be ≥ 0.5 by manipulating $P(+|nE)$. So we want: $\frac{P(E)P(+|E)}{P(+|E)P(E)+P(+|nE)P(nE)} \geq \frac{1}{2}$. We reaarange this inequality as below:

$$\begin{aligned} &\frac{P(E)P(+|E)}{P(+|E)P(E) + P(+|nE)P(nE)} \geq \frac{1}{2} \\ &\Rightarrow 2P(E)P(+|E) \geq P(+|E)P(E) + P(+|nE)P(nE) \\ &\Rightarrow P(E)P(+|E) \geq P(+|nE)P(nE) \\ &\Rightarrow \frac{P(E)P(+|E)}{P(nE)} \geq P(+|nE) \\ &\Rightarrow \frac{\frac{0.95}{300{,}000}}{\frac{299{,}999}{300{,}000}} \geq P(+|nE) \\ &\Rightarrow \frac{19}{5999980} \geq P(+|nE) \end{aligned}$$

So we would want the probability of correctlyt identifying bags NOT containing explosives to be at least $1 - \frac{19}{5999980} \approx 99.9997\%$.

c. Is it possible to obtain a probability of 0.5 of correctly identifying a bag as containing explosives by instead increasing the chance of correctly identifying bags containing explosives? Explain your answer.

Solution:

No. Let the probability of correctly identifying bags containing explosives equal 1, i.e $P(+|E) = 1$, the probability of correctly identifying a bag as containing explosive $P(E|+) = \frac{1\cdot\frac{1}{300{,}000}}{1\cdot\frac{1}{300{,}000}+(1-0.995)\cdot\frac{299{,}999}{300{,}000}} = 6.66 \times 10^{-4}$. Even if $P(+|E) = 1$, $P(E|+)$ is stil very small.

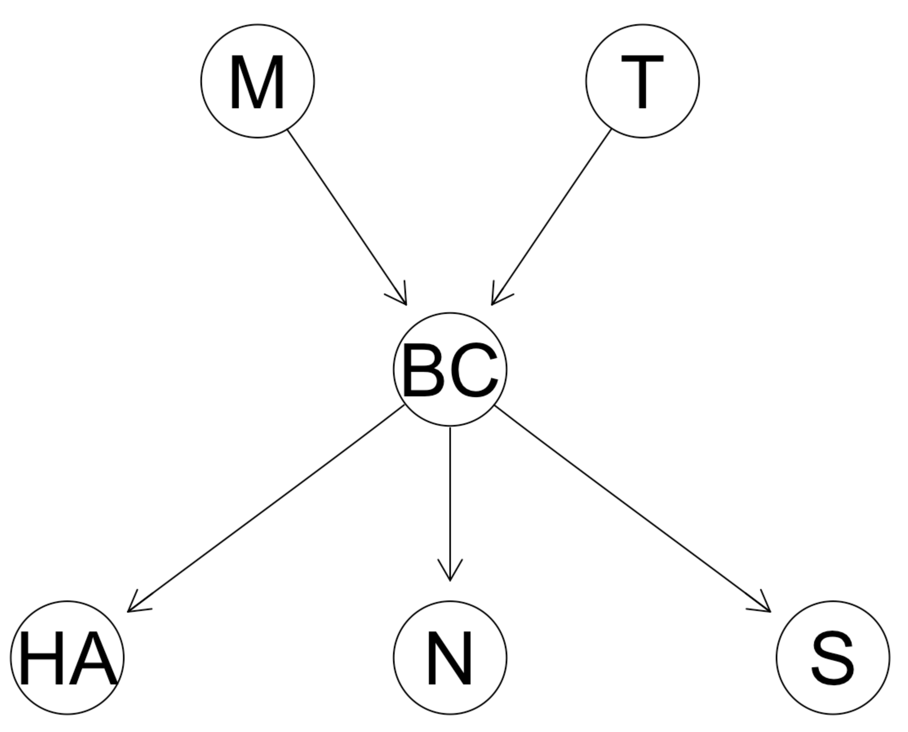
Problem 2

Certain medications and traumas can both cause blood clots. A blood clot can lead to a stroke, heart attack, or it could simply dissolve on its own and have no health implications.

a. Create a DAG that represents this situation.

Solution:

```r
library(Rgraphviz)
library(bnlearn)
dag = model2network("[M][T][BC|M:T][S|BC][HA|BC][N|BC]")
graphviz.plot(dag)
```



b. The following probability information is given where M = medication, T = trauma, BC = blood clot, HA = heart attach, N = nothing, and S = stroke. T stands for true, or this event did occur. F stands for false, or this event did not occur.

P(M=T)	0.2
P(M=F)	0.8
P(T=T)	0.05
P(T=F)	0.95

M	T	P(BC=T)	P(BC=F)
T	T	0.95	0.05
T	F	0.3	0.7
F	T	0.6	0.4
F	F	0.9	0.1

BC	P(HA=T)	P(HA=F)	P(S=T)	P(S=F)	P(N=T)	P(N=F)
T	0.4	0.6	0.35	0.65	0.25	0.75
F	0.15	0.85	0.1	0.9	0.75	0.25

What is the probability that a person will develop a blood clot as a result of both medication and trauma, then have no medical implications?

Solution:

$$\begin{aligned} &P(\mathrm{BC=T,\ N=T|M=T,\ T=T}) \\ &= P(\mathrm{N=T|BC=T,M=T,\ T=T})P(\mathrm{BC=T|M=T,\ T=T}) \\ &= P(\mathrm{N=T|BC=T})P(\mathrm{BC=T|M=T,\ T=T}) \\ &= .75 * .95 = .2375 \end{aligned}$$

Problem 3

Bayesian Network The following Bayesian network represents a joint distribution over the variables: Season(S), Flu(F), Dehydration(D), Chills(C), Headache(H), Nausea(N), and Dizziness(Z). The conditional probability tables for Bayesian network are shown below.

a. What is the probability that you have the flu, given that it is winter, you have a headache, and you know that you are dehydrated?

Solution:

This translates to $P(\mathrm{Flu = true \mid Season = winter, Headache = true, Dehydration = true})$.

$$\begin{aligned} &P(\mathrm{F = true \mid S = winter, H = true, D = true}) \\ &= \frac{P(\mathrm{F = true\ , S = winter, H = true, D = true})}{P(\mathrm{\ S = winter, H = true, D = true})} \\ &= \frac{P(\mathrm{F = true\ , S = winter, H = true, D = true})}{\sum_f P(\mathrm{Flu = f\ ,S = winter, H = true, D = true})} \\ &= \frac{P(\mathrm{H = true \mid F = true, D = true})P(\mathrm{F = true \mid S = winter})P(\mathrm{D = true \mid S = winter})P(\mathrm{S = winter})}{\sum_f P(\mathrm{H = true \mid F = f, D = true})P(\mathrm{F = f \mid S = winter})P(\mathrm{D = true \mid S = winter})P(\mathrm{S = winter})} \\ &= \frac{0.9\cdot 0.4\cdot 0.1\cdot 0.5}{0.9\cdot 0.4\cdot 0.1\cdot 0.5 + 0.8\cdot 0.6\cdot 0.1\cdot 0.5} = \frac{0.018}{0.018+0.024} = 0.43 \end{aligned}$$

b. Does knowing you are dehydrated increase or decrease your probability of having the flu? Intuitively, does this make sense?

$$\begin{aligned} P(\mathrm{F = true|D = true}) &= \frac{P(\mathrm{F = ture, D = true})}{\sum_f P(\mathrm{F = f, D = true})} \\ &= \frac{\sum_s P(\mathrm{F = true, D = true|\ S = s})P(\mathrm{S = s})}{\sum_f \sum_s P(\mathrm{F = f, D = true|\ S = s})P(\mathrm{S = s})} \\ &= \frac{\sum_s P(\mathrm{F = true|S = s})P(\mathrm{D = true|S = s})P(S = s)}{\sum_f \sum_s P(\mathrm{F = f|S = s})P(\mathrm{D = true|S = s})P(S = s)} \\ &= \frac{0.1*0.3*0.5+0.4*0.1*0.5}{0.1*0.3*0.5+0.4*0.1*0.5+0.9*0.3*0.5+0.6*0.3*0.5} = 0.175 \end{aligned}$$

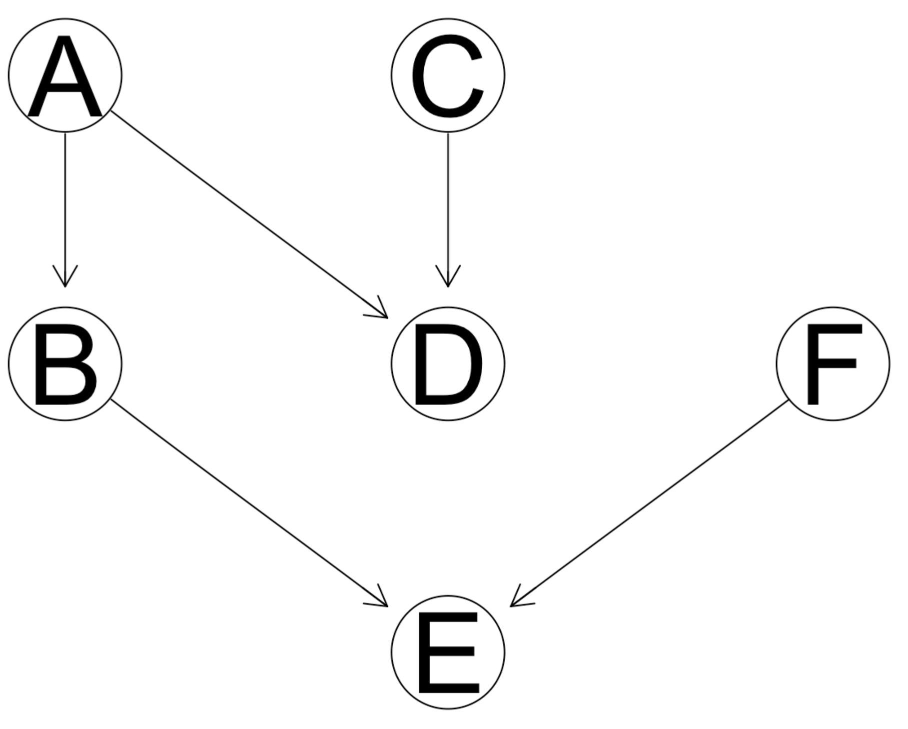
$$P(\mathrm{F = true}) = \sum_s P(\mathrm{F = true}|S = s)P(S = s) = 0.4*0.5+0.1*0.5 = 0.25.$$

Knowing the that you are dehydrated decrease the likelihood that you have the flu. This makes sense because the headache symptom is "explained away" by dehydration.

Problem 4

Write R codes to recover the following Bayesian network and answer the following questions. Explan your answer and write R codes to verify if your reasoning is correct.

```r
# install.packages("BiocManager")
# BiocManager::install("Rgraphviz")
library(Rgraphviz)
library(bnlearn)
# load the data.
data(learning.test)
# create and plot the network structure.
dag = model2network("[A][C][F][B|A][D|A:C][E|B:F]")
graphviz.plot(dag)
```




```r
dsep(dag, 'A','F','B')
```

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

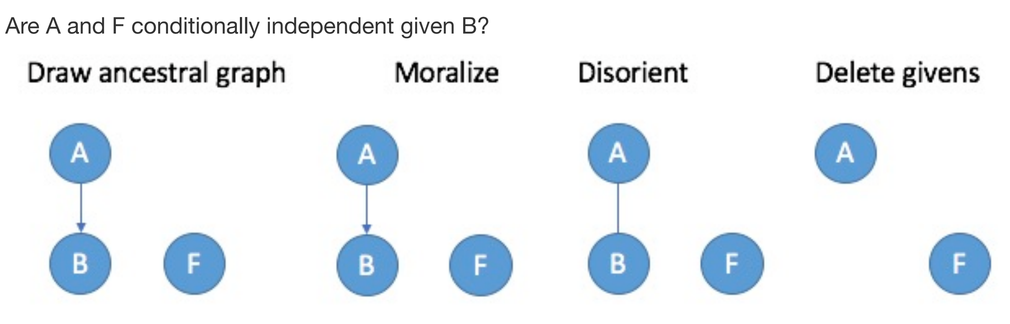
```r
dsep(dag, 'A','F','E')
```

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

```r
dsep(dag, 'A','F')
```

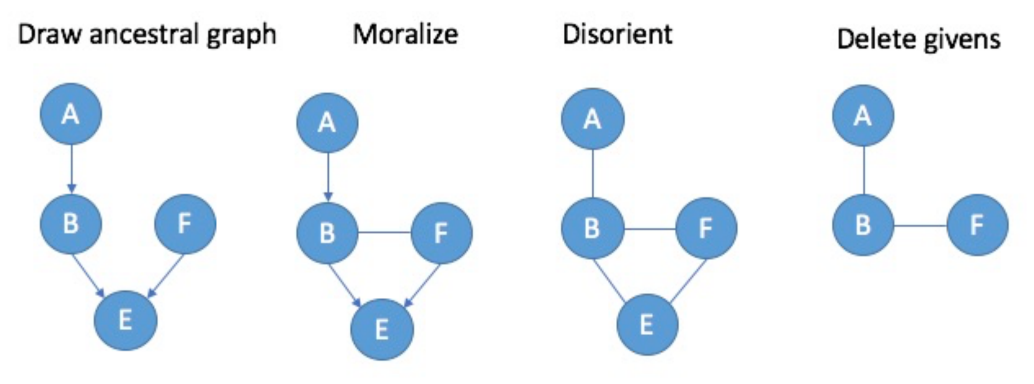
```
## [1] TRUE
```

a. Are A and F conditionally independent given B?



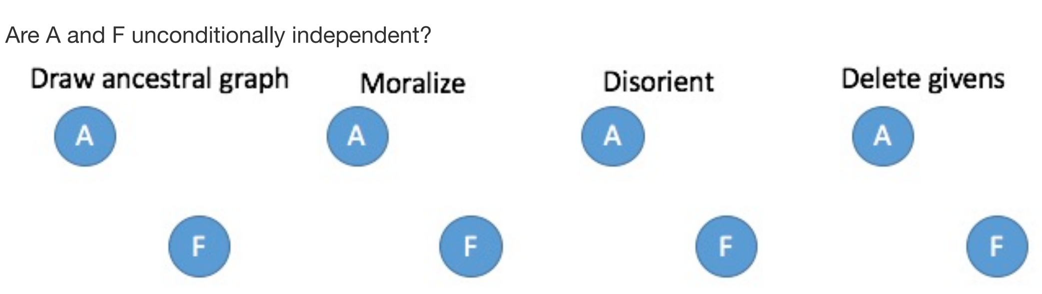
Yes, A and F are not connected.

b. Are A and F conditionally independent given E?



No, A and F are connected through B.

c. Are A and F unconditionally independent?



Yes, A and F are not connected.