TDT4171 Artificial Intelligence Methods

Assignment 1

Written Spring 2022 by Ø. Solbø

Problem 1: Probability

Problem 1.1

Using that X is a random variable representing number of bananas eaten.

Task a)

$$\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) \\
&= 1 - (P(X = 2) + P(X = 1) + P(X = 0)) \\
&= 1 - (0.03 + 0.18 + 0.24) \\
&= 0.55
\end{aligned}$$

Task 1b)

$$\begin{aligned}
P(X \leq 4) &= 1 - P(X \geq 5) \\
&= 1 - 0.17 \\
&= 0.83
\end{aligned}$$

Task 1c)

$$\begin{aligned}
P(X \geq 4) &= P(X = 4) + P(X \geq 5) \\
&= 0.1 + 0.17 \\
&= 0.27
\end{aligned}$$

Problem 1.2

Defining the random variables R and H, which represents number of rotten and healthy apples respectively. Note that $\bar{R} = H$, as they are complement of each other.

Using the theorem of total probability, the probability for drawing two healthy apples is

$$\begin{aligned}
P(H = 2) &= P(H = 2|R = 0)P(R = 0) + P(H = 2|R = 1)P(R = 1) + P(H = 2|R = 2)P(R = 2) \\
&= 0.6 + \frac{19}{20}\frac{18}{19}0.3 + \frac{18}{20}\frac{17}{19}0.1 \\
&= 0.9505
\end{aligned}$$

Task a)

$$\begin{aligned}
P(R = 0|H = 2) &= \frac{P(H = 2|R = 0)P(R = 0)}{P(H = 2)} \\
&= \frac{P(R = 0)}{P(H = 2)} \\
&= \frac{0.6}{0.9505} \\
&= 0.631
\end{aligned}$$

In the second equality, it is used that $P(H=2|R=0)=1$. Verbally: given that there are zero rotten apples, you are guaranteed to draw a healthy apple.

Task 1b)

$$
\begin{aligned}
P(R=1|H=2) &= \frac{P(H=2|R=1)P(R=1)}{P(H=2)} \\
&= \frac{19}{20}\frac{18}{19}\frac{0.3}{0.9505} \\
&= 0.284
\end{aligned}
$$

Task 1c)

$$
\begin{aligned}
P(R=2|H=2) &= \frac{P(H=2|R=2)P(R=2)}{P(H=2)} \\
&= \frac{18}{20}\frac{17}{19}\frac{0.1}{0.9505} \\
&= 0.085
\end{aligned}
$$

Problem 1.3

Defining the random variables S and N, which represents that the person is sick and testing negative respectively. Using the information given in the task, the following probabilities could be determined:

$$
\begin{aligned}
P(S) &= 0.07 \\
P(N|\bar{S}) &= 0.05 \\
P(\bar{N}|S) &= 0.1
\end{aligned}
$$

By using Bayes rule with total probability, the likelihood of having the disease given a negative test is given as

$$
\begin{aligned}
P(S|N) &= \frac{P(N|S)P(S)}{P(N)} \\
&= \frac{P(N|S)P(S)}{P(N|S)P(S)+P(N|\bar{S})P(\bar{S})} \\
&= \frac{(1-P(\bar{N}|S))P(S)}{(1-P(\bar{N}|S))P(S)+P(N|\bar{S})P(\bar{S})} \\
&= \frac{0.9\cdot 0.07}{0.9\cdot 0.07+0.05\cdot 0.93} \\
&= 0.5753
\end{aligned}
$$

This probability seems far too high, given that the test has an acceptable negative and positive predicitive value. With a low prevalence of disease in the population, one ought to expect the probability of disease to be lower given a negative test. This indicates that I have done a mistake somewhere.

Problem 1.4

Using X as a variable to indicate number of defective sets. Number of possible combinations, $N(\cdot)$ ending with at least 2 defective sets given as

$$
\begin{aligned}
N(X\geq 2) &= N(X=2)+N(X=3) \\
&= 2\cdot 1\cdot 9\cdot 8\cdot 7+3\cdot 2\cdot 1\cdot 9\cdot 8 \\
&= 1008+432 \\
&= 1440
\end{aligned}
$$

Problem 2: Bayesian Network Construction

Understanding of the system

The real world is a complex system, and any model cannot include all of the nonlinearities between the factors. Thus, in reality there will be a dependency between all of the variables, however it may be difficult to assess the strength of the dependencies. Some of these may also be affected by other non-modelled variables such as culture. It can therefore be difficult to assess which variables affect each other directly, whereas many will be dually connected in reality. For example having a larger income makes one more likely to eat healthy and exercise, reducing probability of disease. However if a disease does occur, house income could be reduced due to sick leave. In such cases, I will prioritize the one with the strongest connection/dependency.

To 'quickly' summarize my understanding of the system variables, and how I (partly) think they condition each other:

- From medical studies, we know that the prevalence of disease is strongly linked to the history of disease. If either your family or yourself have had a lot of disease, you are more likely to be sick.
- From medical studies, there is a correlation between an healthy diet and disease. An healthy diet will limit the amount of alcohol as well as eating a variety of food.
- In reality, having a healthy diet should induce more varied food and less alchohol. Thus, one should expect people eating fish also to eat fiber and vice versa. In this system, it is difficult to assess how the diets affect each other. I am uncertain how to accurately decide the conditioning, as the system is actually more complex, and a bayesian network might no be sufficient to represent this part of the system.
- The correlation between a healthy diet and prevalence of disease, is likely also affected by the amount of exercise. This will be a hidden variable in this system. One could argue that if one can afford a healthy diet, exercise and being able to do medical checkups, one is less likely to suffer a major illness. This means that the diet (fish, fiber, drinking) and illness conditions on house income.
- Statistically speaking, religion is more likely to be connected to lower house income, however in this context, it is more likely to be related to the culture. Therefore there might be another hidden variable affecting both religion and household economy.
- Having both parents working, will give more income to the household, however it will also affect how much time and energy the parents have for children. Raising children does also take a lot of time and energy. If both parents are working, they might have prioritized their work/studies over having children.
- As some religions have strict rules/conditions on diet, religion affects both drinking habits and diet.
- Number of children will be conditioned by working parents, house income and history of illness. Working parents deduced above. For house income; having children is expensive, however it could be a viable way to get more working hands for poorer households. For richer families, having fewer children is a viable strategy to maintain control over finances. History of illness might also affect how willing or how easy it is to get children. Genetical mutations might put somebody off, while other might have problem concieving due to earlier cancer treatment.
- The argument could be made of a change in diet (and drinking) due to number of children. Especially younger children could have a dislike in fish, affecting how much fish the family eats. For drinking, one should hope that the parents will drink less when having children around.
- Having more children will expose the household to a larger part of the community. Especially now under covid, having children in school, has been a major underlining reason for spreading

covid.

Abbreviations

Different abbreviations used in the diagram:

- I: Illness
- HI: History of illness
- D: Drinking-habits
- Fish: Fish-eating habits
- Fiber: Fiber-eating habits
- R: Religion
- HE: House economy (house income)
- WP: Working parents
- NC: (not connected) Number of children

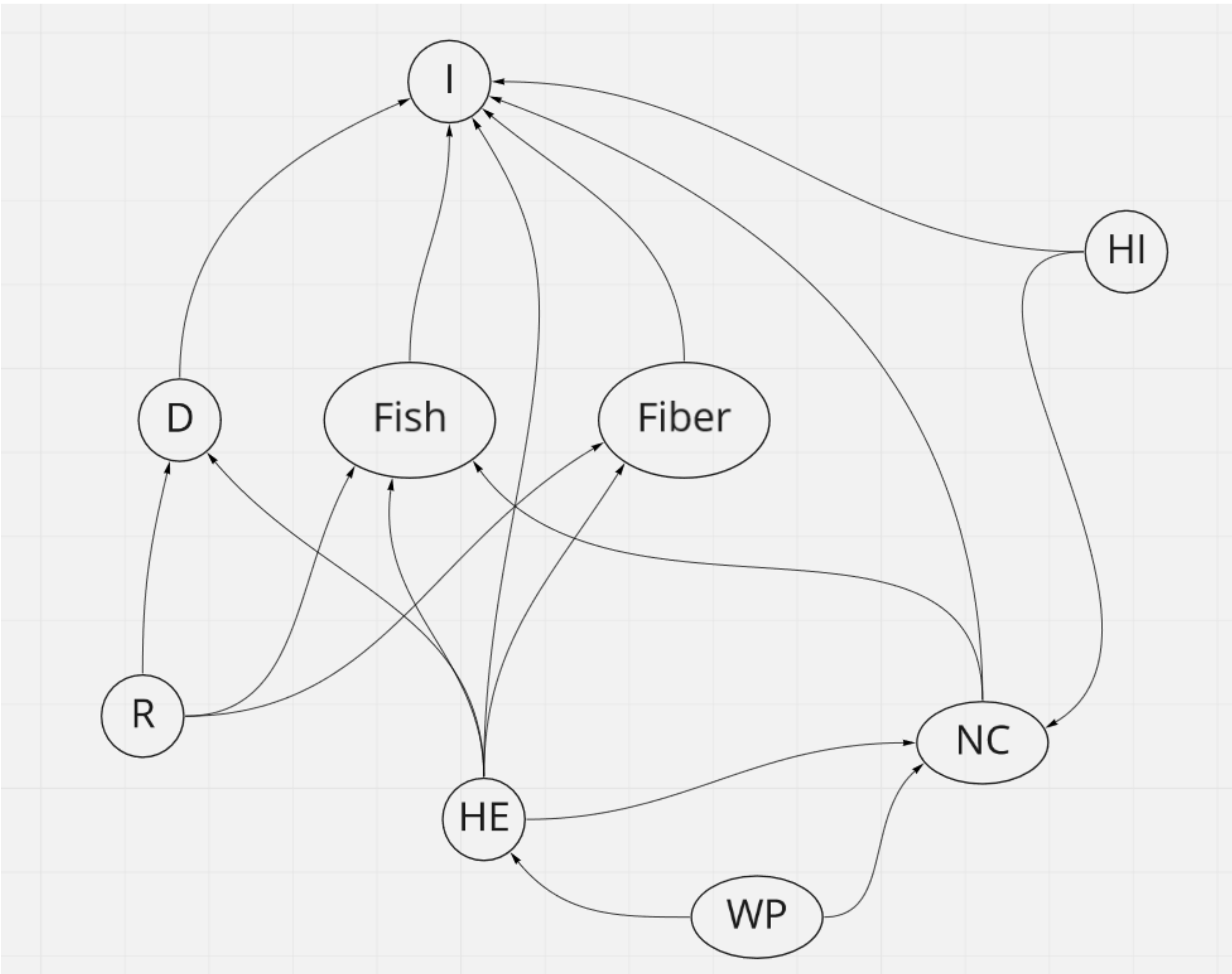

Figure 1: Bayes net created for task 1.2

Independent variables

Based on the figure, I have set the variables WP, R and HI as independent.

Religion is assumed to be determined more on a cultural reasoning. If the society is more religious, independent on how large the income is and how educated people are, children are more likely of being indoctrinated at an early age. Thus continouing the religious pressure in the future. In other words assumed dependent only on culture and not any of the variables used in the Bayes Network.

I am a bit unsure about WP as independent, because the argument could be made that it is also determined by the house income. If the house income is too low, both parents must work (possible multiple jobs) to get things to go around. In reality however, WP is therefore dependent of HE. However, I think that the pressure from society/culture of having both parents working must also be taking into consideration. The value that each parents bring into the house income is therefore more important when creating a bayes net.

HI is set as independent, however this will be affected by current disease, diet and drinking habits, as well as exposure to different diseases/causes. However I think that current illness is more dependent on history of illness, because history of illness will affect f.ex. antibodies and the reaction to treatment or illness.

Problem 3: The Monty Hall problem

The bayesian network have three nodes, where the node OpenedByOfficial will depend on ContainsPrize and MyChoice. The bayesian network is shown in figure 2.

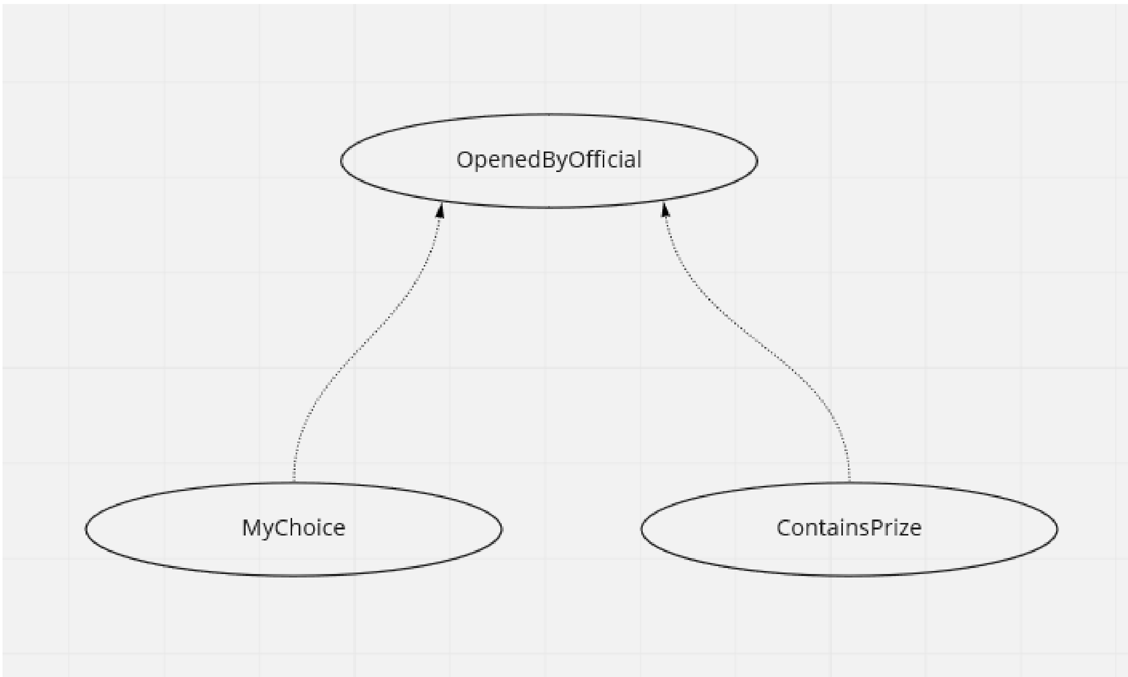

Figure 2: Bayes net created for the Monty Hall problem in task 1.3

I assume that the probability tables should reflect the probability of which door the official will open. The probability-tables are given in tables 1 to 3. Warning: This is the second version of the probability-tables. With the first version, I forgot to normalize, leaving the cumulative probability of the table to exceed 1. In the first version I calculated the probability of rows (official's choise of door) dependent on columns (my choice of doors). However, that does not give a correct probability table.

Prize behind: Door 0	**My choice**		
Opened by official	Door 0	Door 1	Door 2
Door 0	0	0	0
Door 1	$\frac{1}{6}$	0	$\frac{1}{3}$
Door 2	$\frac{1}{6}$	$\frac{1}{3}$	0

Table 1: Probability-table showing the probability for which door the official will open, given that the prize is behind door 0.

Prize behind: Door 1	**My choice**		
Opened by official	Door 0	Door 1	Door 2
Door 0	0	$\frac{1}{6}$	$\frac{1}{3}$
Door 1	0	0	0
Door 2	$\frac{1}{3}$	$\frac{1}{6}$	0

Table 2: Probability-table showing the probability for which door the official will open, given that the prize is behind door 1.

Prize behind: Door 2	**My choice**		
Opened by official	Door 0	Door 1	Door 2
Door 0	0	$\frac{1}{3}$	$\frac{1}{6}$
Door 1	$\frac{1}{3}$	0	$\frac{1}{6}$
Door 2	0	0	0

Table 3: Probability-table showing the probability for which door the official will open, given that the prize is behind door 2.

The key behind deriving the probability tables above, is that the official knows where the prize is located and is forced to open a door without a prize. In two of three cases, where we do not originally pick the door containing a prize, the official is forced to open an empty door. This means that in two of three cases, one expect to win by switching doors. In other words; yes, you should switch doors to increase the expected value of winning.

References