Exercise 1

Problem 1

The probability that a person has 0, 1, 2, 3, 4, or 5 or more siblings is 0.15, 0.49, 0.27, 0.06, 0.02, 0.01, respectively.

a) *What is the probability that a child has at most 2 siblings?* ***0.91***

We know that the number of siblings a person has can be represented as discrete variables. As it is impossible for a person to have both X and Y siblings at the same time, with X ≠ Y, and it is impossible for a person to have a number of siblings not represented in the range [0, ∞), the variables are mutually exclusive and exhaustive. We thus have:

P(at most 2 siblings) = P(0 siblings) + P(1 sibling) + P(2 siblings)

= 0.15 + 0.49 + 0.27

= 0.91

We can check this, as

P(at most 2 siblings) = 1 – (P(3 siblings) + P(4 siblings) + P(5 or more siblings))

= 1 – (0.06 + 0.02 + 0.01)

= 1 – 0.09 = 0.91

b) *What is the probability that a child has more than 2 siblings given that he has at least 1 sibling?* ***<u>0.11</u>***

We know from above that the number of siblings can be represented as mutually exclusive, exhaustive variables. We can see that:

P(more than 2 siblings | at least 1 sibling)

$$= \frac{P(more\ than\ 2\ siblings\ \wedge at\ least\ 1\ sibling)}{P(at\ least\ 1\ sibling)}$$

A person can only have 2 or more siblings if they have at least 1 sibling, meaning that $P(more\ than\ 2\ siblings\ \wedge at\ least\ 1\ sibling) = P(more\ than\ 2\ siblings)$. We also know that $P(more\ than\ 2\ siblings) = 1 - P(at\ most\ 2\ siblings) = 0.09$. We thus have:

P(more than 2 siblings | at least 1 sibling)

$$= \frac{P(more\ than\ 2\ siblings)}{P(at\ least\ 1\ sibling)}$$

$$= \frac{1 - P(at\ most\ 2\ siblings)}{1 - P(0\ siblings)}$$

$$= \frac{0.09}{0.85} \approx \mathbf{0.11}$$

c) *Three friends who are not siblings are gathered. What is the probability that they combined have three siblings?* ***<u>0.24</u>***

As the three friends are not siblings, we can safely assume that their number of siblings are independent variables (there are several ways this can be untrue, such as their families knowing each other, coming from the same area or social backgrounds, but we will let the assumption stand). The following rule is thus true:

P(A, B) = P(A) * P(B)

We look at the different combinations that adds up to three siblings in total:

Person A	Person B	Person C	P(A)	P(B)	P(C)	P(A)*P(B)*P(C)
3	0	0	0.06	0.15	0.15	0.00135
2	1	0	0.27	0.49	0.15	0.019845
2	0	1	0.27	0.15	0.49	0.019845
1	2	0	0.49	0.27	0.15	0.019845
1	1	1	0.49	0.49	0.49	0.117649
1	0	2	0.49	0.15	0.27	0.019845
0	3	0	0.15	0.06	0.15	0.00135
0	2	1	0.15	0.27	0.49	0.019845
0	1	2	0.15	0.49	0.27	0.019845
0	0	3	0.15	0.15	0.06	0.00135
Sum						0.240769

<u>We see that the sum of the probability of all combinations is 0.24</u>

d) *Emma and Jacob are not siblings, but combined they have a total of 3 siblings. What is the probability that Emma has no siblings?* ***0.318***

The only combination which satisfies both Emma having no siblings and them combined having 3 siblings is Jacob having 3 siblings. We are therefore looking for:

P(Jacob has 3 siblings ∧ Emma has no siblings | combined 3 siblings)

$$= \frac{P(Jacob\ has\ 3\ siblings\ \wedge Emma\ has\ no\ siblings)}{P(combined\ 3\ siblings)}$$

As Jacob and Emma are not siblings, we can assume that their number of siblings are independent variables (see task c). This means that P(A ∧ B) = P(A)*P(B). We look at the different combinations leading to a combined number of 3 siblings, like we did in (c):

Jacob siblings	Emma siblings	P(J = j)	P(E = e)	P(J = j ∧ E = e)
0	3	0.15	0.06	0.009
1	2	0.49	0.27	0.1323
2	1	0.27	0.49	0.1323
3	0	0.06	0.15	0.009
Sum				0.2826

From this table, we can see that P(Jacob has 3 siblings ∧ Emma has no siblings) = 0.009 and P(combined 3 siblings) is 0.2826. We find

$$P\left(Jacob\ has\ 3\ siblings\ \bigwedge Emma\ has\ no\ siblings \middle| combined\ 3\ siblings\right)$$

$$= \frac{P(Jacob\ has\ 3\ siblings\ \wedge Emma\ has\ no\ siblings)}{P(combined\ 3\ siblings)}$$

$$= \frac{0.009}{0.2826} = \mathbf{0.318}$$

Problem 2

Given the Bayesian network structure (in the task), decide whether the statements are true or false. Justify each answer with an explanation.

a) If every variable in the network has a Boolean state, then the Bayesian network can be represented with 18 numbers. ***TRUE***

We know that each node can be represented by 2^p numbers, where p is the number of parents the node has. We can create a table to represent each node and respective needed representation:

Node	**Number of parents**	**Number needed to represent**
A	0	1
B	1	2
C	1	2
D	1	2
E	2	4
F	2	4
G	1	2
H	0	1
Number needed for whole network:		**18**

b) G ⊥⊥ A. ***TRUE***

We know that each node is conditionally independent of non-successors, given its parents. As A has no parents, it is conditionally independent of non-successors given an empty set, meaning it is unconditionally independent of its non-successors. As G is not a successor of A, A ⊥⊥ G. Due to symmetry, G ⊥⊥ A meaning that G and A are independent variables.

c) E ⊥⊥ H | {D, G} ***TRUE***

Looking at H, we can see that its Markov Blanket consists of {D, G}, meaning that H is conditionally independent of any node given those two nodes. As they are given in the task, H is conditionally independent of any node, meaning that it must be conditionally independent of E.

d) *E ⊥⊥ H | {C, D, F}* ***FALSE***

I will answer this question by looking at paths in the Bayesian network. Two nodes X and Y are conditionally independent given a set of nodes S if there does not exists a path between X and Y of nodes that are conditionally dependent on Y. As an example, if the network contains a path Y → S1 → S2 → X, where arrows indicate dependencies, then X and Y are not conditionally independent. This is also called an active path.

With F given, E and G has a "common cause", meaning that E and G are conditionally dependent given F. As H is a direct parent of G, G and H are dependent. Combining these two, we can see that there exists an active path in E → H → G, meaning that E and H are not conditionally independent given {C, D, F}

Problem 3

The Bayesian network [given] contains only binary states. The conditional probability for each state is listed. From the Bayesian network, calculate the following probabilities:

a) P(b)

$$P(b) = \sum_{All\ values\ of\ Parent(b)} P(b \mid Parent(b)) * P(Parent(b))$$

$$= P(b \mid a) * P(a) + P(b \mid \neg a) * P(\neg a)$$

$$= 0.5 * 0.8 + 0.2 * 0.2$$

$$\underline{= \mathbf{0.44}}$$

b) P(d)

Note that if P(b) = 0.44, then P(¬b) = 0.56.

As above, $P(d) = \sum_{All\ values\ of\ Parent(d)} P(d \mid Parent(d)) * P(Parent(d))$

$$= P(d \mid b) * P(b) + P(d \mid \neg b) * P(\neg b)$$

$$= 0.6 * 0.44 + 0.8 * 0.56$$

$$\underline{= \mathbf{0.712}}$$

c) P(c | ¬d)

Bayes' rule gives us $P(b \mid \neg d) = \frac{P(\neg d \mid b) * P(b)}{P(\neg d)}$

$$= \frac{0.4 * 0.44}{1 - 0.712}$$

$$= \frac{0.176}{0.288} = 0.611$$

We can clearly see that $P(\neg b \mid \neg d) = 1 - 0.611 = 0.389$

We can rewrite $P(c \mid \neg d)$ as the sum of two joint probabilities:

$$P(c \mid \neg d)$$

$$= P(c, b \mid \neg d) + P(c, \neg b \mid \neg d)$$

$$= \frac{P(c, b, \neg d)}{P(\neg d)} + \frac{P(c, \neg b, \neg d)}{P(\neg d)}$$

We can apply the chain rule on this, and get

$$= \frac{P(c \mid b, \neg d) * P(b \mid \neg d) * P(\neg d)}{P(\neg d)} + \frac{P(c \mid \neg b, \neg d) * P(\neg b \mid \neg d) * P(\neg d)}{P(\neg d)}$$

As b is the only parent of c, we can reduce the following:

$$P(c \mid b, \neg d) = P(c \mid b) \text{ and } P(c \mid \neg b, \neg d) = P(c \mid \neg b)$$

We then have

$$P(c \mid \neg d)$$
$$= P(c \mid b) * P(b \mid \neg d) + P(c \mid \neg b) * P(\neg b \mid \neg d)$$

We are given the first and third probability in the task and calculated the rest above. We can therefore find that

$$P(c \mid \neg d)$$
$$= 0.1 * 0.611 + 0.3 * 0.389$$
$$= \mathbf{0.178}$$

d) P(a | ¬c, d)

Like in (c), we can rewrite this as the sum of two joint probabilities:

$$P(a \mid \neg c, d)$$
$$= P(a, b \mid \neg c, d) + P(a, \neg b \mid \neg c, d)$$
$$= \frac{P(a, b, \neg c, d)}{P(\neg c, d)} + \frac{P(a, \neg b, \neg c, d)}{P(\neg c, d)}$$

We can apply the chain rule:

$$= \frac{P(a \mid b, \neg c, d) * P(\neg c \mid b, d) * P(d \mid b) * P(b)}{P(\neg c, d)}$$
$$+ \frac{P(a \mid \neg b, \neg c, d) * P(\neg c \mid \neg b, d) * P(d \mid b) * P(b)}{P(\neg c, d)}$$

As b is the only child of a, $P(a \mid b, \neg c, d) = P(a \mid b)$ and $P(a \mid \neg b, \neg c, d) = P(a \mid \neg b)$.

Bayes' rule gives us:

$$P(a \mid b) = \frac{P(b \mid a) * P(a)}{P(b)}$$
$$= \frac{0.5 * 0.8}{0.44} = 0.909$$
$$P(a \mid \neg b) = \frac{P(\neg b \mid a) * P(a)}{P(\neg b)}$$
$$= \frac{0.5 * 0.8}{1 - 0.44} = 0.714$$

We have the same with $P(\neg c \mid b, d) = P(\neg c \mid b)$ and $P(\neg c \mid \neg b, d) = P(\neg c \mid \neg b)$ as well.

We can find

$$
\begin{aligned}
&P(\neg c, d) \\
&= P(\neg c \mid d) * P(d)
\end{aligned}
$$

Using the method from (c):

$$
= [P(\neg c \mid b) * P(b \mid d) + P(\neg c \mid \neg b) * P(\neg b \mid d)] * P(d)
$$

Bayes' rule gives us:

$$
\begin{aligned}
&P(b \mid d) = \frac{P(d \mid b) * P(b)}{P(d)} \\
&= \frac{0.6 * 0.44}{0.712} = 0.371 \\
&P(\neg b \mid d) = 1 - 0.371 = 0.629
\end{aligned}
$$

Thus, we have:

$$
\begin{aligned}
&P(\neg c, d) \\
&= [P(\neg c \mid b) * P(b \mid d) + P(\neg c \mid \neg b) * P(\neg b \mid d)] * P(d) \\
&= [0.9 * 0.371 + 0.7 * 0.629] * 0.712 \\
&= 0.774 * 0.712 = 0.551
\end{aligned}
$$

We combine this, our reductions, and the chain rule:

$$
\begin{aligned}
&P(a \mid \neg c, d) \\
&= \frac{1}{P(\neg c, d)} (P(a \mid b) * P(\neg c \mid b) * P(d \mid b) * P(b) + P(a \mid \neg b) * P(\neg c \mid \neg b) * P(d \mid \neg b) \\
&\qquad * P(\neg b)) \\
&= \frac{1}{0.551} (0.909 * 0.9 * 0.6 * 0.44 + 0.714 * 0.7 * 0.8 * 0.56) \\
&= \frac{1}{0.551} (0.216 + 0.224) = \frac{0.440}{0.551} = \mathbf{0.799}
\end{aligned}
$$

Problem 4

Note that the code runs on the following:

- Python 3.6.9
- Installed pip modules: numpy == 1.19.5

a) *First you need to implement a Bayesian network that supports discrete conditional probability distributions for each state*

See file "Exercise1.py"

b) *Implement the inference by enumeration algorithm found in Figure 14.9 of Artificial Intelligence: A Modern Approach*

See file "Exercise1.py"

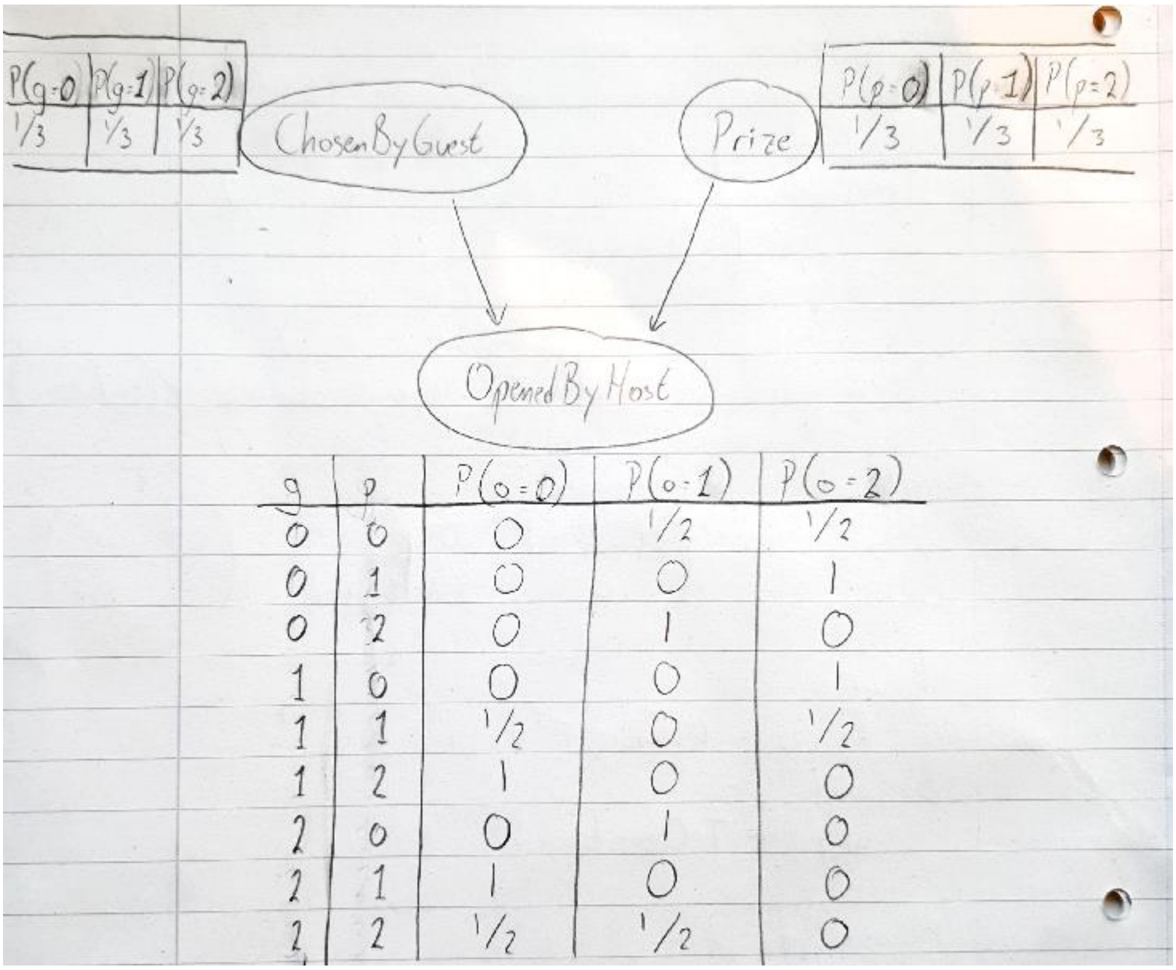
c) *Model the Monty Hall problem as a Bayesian Network using the following states ChosenByGuest, OpenedByHost, and Prize. Use your implementation of inference by enumeration, and the evidence described in the problem statement to answer the question; is it to your advantage to switch your choice? Turn in your code, a drawing of the Bayesian network with conditional probability tables, and the calculated posterior probabilities.*

Note: For simplicity, I have decided to label doors 0-2, instead of 1-3 (as given in the task). This has no effect on the answer, but it matches the internal state representation in my code.

It is advantageous to switch choice. As every combination of ChosenByGuest and OpenedByHost gives the same probability, it is enough to look at the one given in the assignment – P(Prize | ChosenByGuest = 0, OpenedByHost = 2). As we can see from the figure below, the probability of the prize being at door 0 (which the guest selected) is 1/3, while the probability of it being at door 1 (which the guest can switch to) is 2/3. This means that the chance of winning doubles if the guest switches door, no matter the initial selection.

```
Probability distribution, P(Prize | ChosenByGuest = 0, OpenedByHost = 2)
+----------+----------+
| Prize(0) |  0.3333  |
+----------+----------+
| Prize(1) |  0.6667  |
+----------+----------+
| Prize(2) |  0.0000  |
+----------+----------+
```

For simplicity, I have labeled Prize as *p*, ChosenByGuest as *g* and OpenedByHost as *o* in the conditional probability tables below.



P(g=0)	P(g=1)	P(g=2)
1/3	1/3	1/3

P(p=0)	P(p=1)	P(p=2)
1/3	1/3	1/3

g	p	P(o=0)	P(o=1)	P(o=2)
0	0	0	1/2	1/2
0	1	0	0	1
0	2	0	1	0
1	0	0	0	1
1	1	1/2	0	1/2
1	2	1	0	0
2	0	0	1	0
2	1	1	0	0
2	2	1/2	1/2	0