ROB311 - TD5 - Bayesian Network

BRAMBILLA Davide Luigi - GOMES DA SILVA Rafael

Octobre 14, 2019

Introduction

A **Bayesian network** is a graphical model where it is possible to represent a set of variables and their dependencies through a ***D**irected **A**cyclic **G**raph(DAG).*

Their aim is to model conditional dependence and causation by inserting in the graph arrows that link the different elements.

In particular, inside the graph, we have that the *nodes* are the variables and that the *arrows* are the dependencies between variables. Furthermore the absence of an arrow means the independence between the nodes.

Each node has a proper conditional probability $P(X_i|Parents(X_i))$ that quantifies the dependence of the parents on the current node.

Q1 - The Bayesian Network of the problem

The problem proposed models the system used by a hospital to help in the diagnostic of *tuberculosis* (T), *lung cancer* (C) and *bronchitis* (B).

The two possible causes that might cause those diseases in a person are: the fact that a person has travelled to Asia (A) and the fact that a person is a smoker(S).

Finally we have only two possible tests proposed by the doctors in order to detect the illnesses: the *stethoscope* (St) or the *X-Rays* (X).

The Figure 1 shows the **Bayesian network** that models the problem and its dependencies.

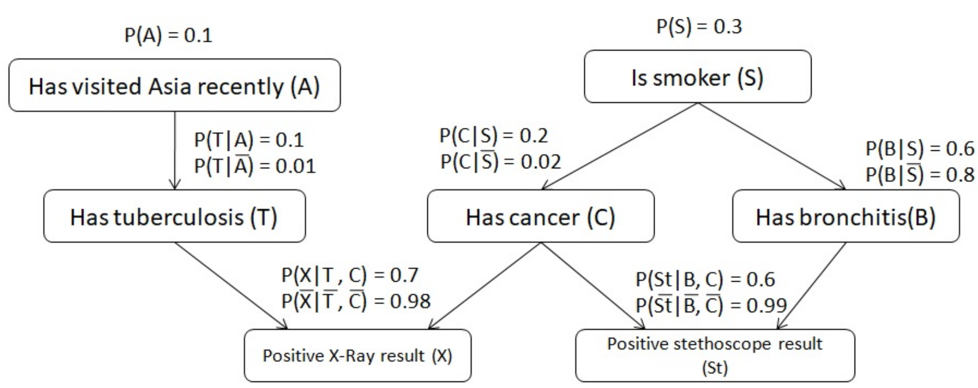

Figure 1: The Bayesian Network

Q2 - The probabilities of the illnesses

Given the model of the Figure 1, we have to consider the case when the patient hasn't travelled to Asia and he or she is not a smoker. For this person, we have to calculate the possibility that he or she has *tuberculosis* or that he or she has *lung cancer* or *bronchitis.*

We will calculate the three probabilities:

- $p(T \mid \bar{A}, \bar{S})$: to have *tuberculosis* without having traveled to Asia($\bar{A}$) **and** not being a smoker($\bar{S}$).
- $p(C \mid \bar{A}, \bar{S})$: to have *lust cancer* without having traveled to Asia($\bar{A}$) **and** not being a smoker($\bar{S}$).
- $p(B \mid \bar{A}, \bar{S})$: to have *bronchitis* without having traveled to Asia($\bar{A}$) **and** not being a smoker($\bar{S}$).

Since the three probabilities are *conditional* we are going to use the Bayesian formula in order to asses them.

$$p(T \mid \bar{A}, \bar{S}) = \frac{p(T, \bar{A}, \bar{S})}{p(\bar{A}, \bar{S})} = \frac{p(T, \bar{S}, \bar{A})}{p(\bar{S}) \cdot p(\bar{A})} = \frac{p(\bar{S}) \cdot p(T, \bar{A})}{p(\bar{S}) \cdot p(\bar{A})} = \frac{p(T, \bar{A})}{p(\bar{A})} = p(T|\bar{A}) = 0.01$$

Since we have that the event $\bar{S}$ and the events $T\ and\ \bar{A}$ are independents we can separate the probability of the three events at the same time with the product between the probability of $\bar{S}$ and the probability of $T\ and\ \bar{A}$.

The other two probabilities are obtained in the same way and are:

$$p(C \mid \bar{A}, \bar{S}) = \frac{p(C, \bar{A}, \bar{S})}{p(\bar{A}, \bar{S})} = \frac{p(C, \bar{A}, \bar{S})}{p(\bar{A}) \cdot p(\bar{S})} = \frac{p(\bar{A}) \cdot p(C, \bar{S})}{p(\bar{A}) \cdot p(\bar{S})} = \frac{p(C, \bar{S})}{p(\bar{S})} = p(C|\bar{S}) = 0.02$$

$$p(B \mid \bar{A}, \bar{S}) = \frac{p(B, \bar{A}, \bar{S})}{p(\bar{A}, \bar{S})} = \frac{p(B, \bar{A}, \bar{S})}{p(\bar{A}) \cdot p(\bar{S})} = \frac{p(\bar{A}) \cdot p(B, \bar{S})}{p(\bar{A}) \cdot p(\bar{S})} = \frac{p(B, \bar{S})}{p(\bar{S})} = p(B|\bar{S}) = 0.8$$

The values obtained are:

$$\begin{aligned} p(T \mid \bar{A}, \bar{S}) &= 0.01 = 1\% \\ p(C \mid \bar{A}, \bar{S}) &= 0.02 = 2\% \\ p(B \mid \bar{A}, \bar{S}) &= 0.8 = 80\% \end{aligned}$$

Finally, we can say that the most probable illness for a person that hasn't traveled to Asia and that isn't a smoker is *bronchitis*.

Q3

Q3.1 Why the doctor decides to auscultate the patient's lung with a stethoscope?

From Q2 we've calculated that the most probable illness is the *bronchitis*, with a probability of 80% of having the disease. However, there is still a slightly chance of the patient having lung cancer (2%). Because of that, the doctor chooses to auscultate the patient's lung with a stethoscope so the probability to effectively detect the disease will be higher. In the end, if the patient is affected by *bronchitis*, the doctor can detect it with an high level of certainty.

Q3.2 The stethoscope test is negative. What is the new inferred diagnosis?

In order to infer the disease that the patient has, knowing that he or she is not a smoker and hasn't been to Asia, and also that the stethoscope result was negative, we used part of the Bayesian network that models the problem (Figure 1) with more details about all the probabilities of each node, as shown in Figure 2.

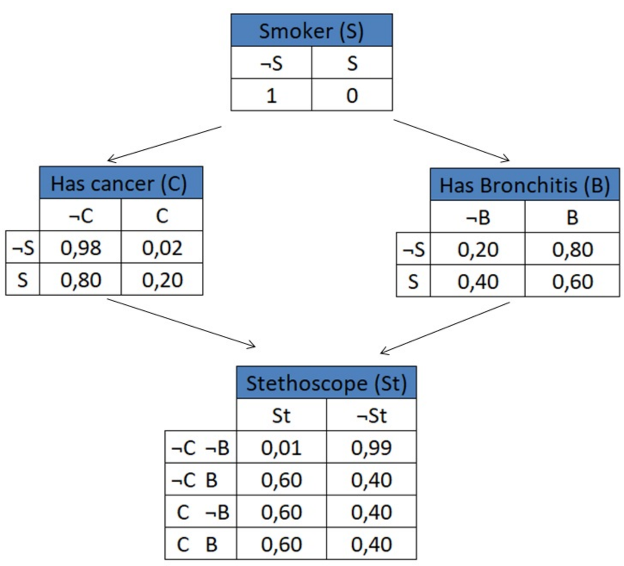

Figure 2: Table diagram used for the inference for a negative stethoscope test

Since the information on the *Stethoscope* test doesn't give any information on the *tuberculosis disease*, its probability will be the same calculated in Q2 (1%).

The probabilities to have *lung cancer* or *bronchitis* disease will instead decrease.
In order to calculate them we have to consider the new information on the *Stethoscope* test that has resulted **negative**.
In particular we will have that:

$$p(B \mid \bar{S}t) = \frac{p(\bar{S}t \mid B)\, p(B)}{p(\bar{S}t)}$$

and

$$p(C \mid \bar{S}t) = \frac{p(\bar{S}t \mid C)\, p(C)}{p(\bar{S}t)}$$

In order to calculate these two probabilities we have to calculate $p(\bar{S}t)$.
In particular we will have that:

$$\begin{aligned} p(\bar{S}t) = {} & p(\bar{S}t|\bar{B}, \bar{C}) \cdot p(\bar{C}|\bar{S}) \cdot p(\bar{B}|\bar{S}) + (\bar{S}t|\bar{B}, C) \cdot p(C|\bar{S}) \cdot p(\bar{B}|\bar{S}) + \\ & p(\bar{S}t|B, \bar{C}) \cdot p(\bar{C}|\bar{S}) \cdot p(B|\bar{S}) + p(\bar{S}t|B, C) \cdot p(C|\bar{S}) \cdot p(B|\bar{S}) \end{aligned}$$

that will have the value of 0.51564(51.56%).
In the end, by reading the values from the tables, we will have that:

$$p(\bar{S}t \mid B) = 0.4$$
$$p(B) = 0.8$$
$$p(\bar{S}t \mid C) = 0.4$$
$$p(C) = 0.02$$

By substituting the values reported above in the previous equations we will have that:

$$p(B \mid \bar{S}t) = 0.6206 = 62.06\%$$
$$p(C \mid \bar{S}t) = 0.155 = 1.55\%$$

For the sake of completeness, we will report here the value calculated before for the *tuberculosis* disease that is still 0.01(1%). So in the end, we can conclude that, even if the *Stethoscope* test is negative the highest probability is still the one related to *bronchitis*.

Q4 - What's the new inferred diagnosis after that the X-Ray test is positive?

In this case we will have to consider that the *X-Ray* test has given a positive response. So we expect the probabilities that the patient has *tuberculosis* or *lung cancer* to be more consistent than in the previous cases.
In the figure 3 below we have reported the left part of our graph with the knowledge that the patient hasn't been to Asia and isn't a smoker.

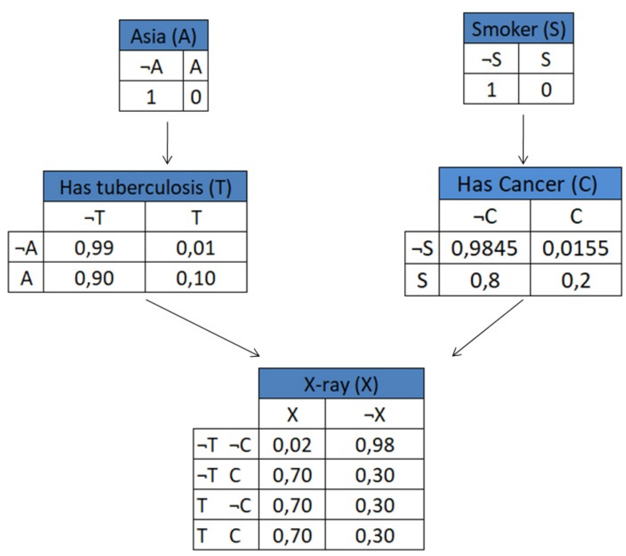

Figure 3: Table diagram used for the inference for a positive X-Ray test

In order to calculate the two probabilities we ave to consider the new informations on the *X-Ray* test that has resulted **positive**.

In particular we will have that:

$$p(C \mid X) = \frac{p(X \mid C)\, p(C)}{p(X)}$$

and

$$p(T \mid X) = \frac{p(X \mid T)\, p(T)}{p(X)}$$

In order to calculate these two probabilities we have to calculate $p(X)$.

In particular we will have that:

$$p(X) = p(X|\bar{T},\bar{C}) \cdot p(\bar{C}|\bar{S}) \cdot p(\bar{T}|\bar{A}) + (X|\bar{T},C) \cdot p(C|\bar{S}) \cdot p(\bar{T}|\bar{A}) + p(X|T,\bar{C}) \cdot p(\bar{C}|\bar{S}) \cdot p(T|\bar{A}) + p(X|T,C) \cdot p(C|\bar{S}) \cdot p(T|\bar{A})$$

that will have the value of 0.0372(3.72%).

In the end, by reading the values from the tables, we will have that:

$$p(X \mid C) = 0.7$$
$$p(C) = 0.0155 \ (\textit{without considering the Stethoscope test would be } 0.02)$$

$$p(X \mid T) = 0.7$$
$$p(T) = 0.01$$

By substituting the values reported above in the previous equations we will have that:

$$p(C \mid X) = 0.2916 = 29.16\%$$
$$p(T \mid X) = 0.1879 = 18.79\%$$

So in the end, we can conclude that, even if the *X-Ray* test is negative the highest probability is still the one related to *bronchitis*.

Q5 - Was the X-Ray needed?

According to the results obtained with the Bayesian network we could conclude that the x-ray was **not** needed. After the positive X-Ray test the percentages of the patient to suffer from *tuberculosis* or from *lung cancer* are still smaller than the one obtained in the *Q3* for a patient that suffers from *bronchitis*, as we can see in the Table 1.

This can be explained by the fact that the knowledge that the patient hasn't been in Asia and that he isn't a smoker give a strong *prior* over the whole problem.

Table 1: Comparison of the results of the inferences obtained for a negative stethoscope test and for a positive X-Ray test

Inference	**Value (%)**
P(B $\mid\neg$St)	62,06
P(C $\mid\neg$St)	1,55
P(C $\mid X$)	29,16
P(T $\mid X$)	18,79