ROB311 - TP5 - Bayes Networks

Brice Tayart

October 18, 2019

Medical diagnosis

A hospital uses a support system for detecting lung problems. The system is designed to help in the diagnosis of tuberculosis, cancer, and bronchitis. The system will use previous data from the hospital gathered from previous consultations.
At the registration in the hospital, a new patient it is asked to fill in a questionnaire and answer 2 questions: "Have you recently visited Asia? " and " Are you a smoker? ".

The data shows that:

- Among all the patients, 10% have recently visited Asia, and 30% are smokers;
- Tuberculosis is present in Asia, and a patient who recently visited Asia has 10% of having tuberculosis and a patient who have not been recently to Asia has only 1% of having tuberculosis;
- Patients that smoke and complain of lung problems have 20% of having cancer (against only 2% for patients that do not smoke);
- Patients that do not smoke are suffering in 80% of cases of only a bronchitis(against only 60% for people that smoke).

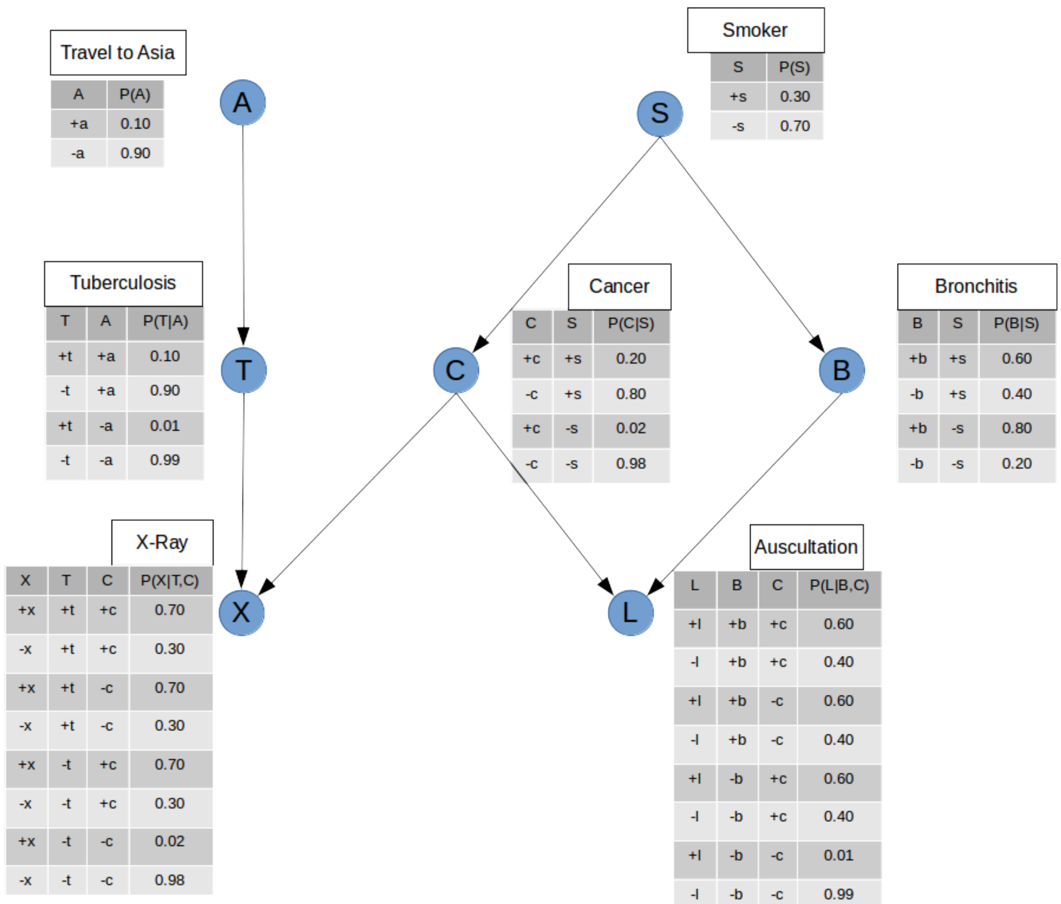
The doctor proposes only 2 tests:

- The doctor auscultates the patient's lungs with a stethoscope. A bronchitis or a lung cancer can be detected in 60% of cases. When the patient has none of these two diseases, the doctor will detect it with a probability of 99%;
- The doctor orders an X-Ray. With the X-Ray the tuberculosis or lung cancer are detected in 70% of cases. If the patient has none of these two diseases, nothing will be observed on the X-Ray with a probability of 98%.

Questions

Question 1

Model this problem using a Bayesian network



A	P(A)
+a	0.10
-a	0.90

S	P(S)
+s	0.30
-s	0.70

T	A	P(T|A)
+t	+a	0.10
-t	+a	0.90
+t	-a	0.01
-t	-a	0.99

C	S	P(C|S)
+c	+s	0.20
-c	+s	0.80
+c	-s	0.02
-c	-s	0.98

B	S	P(B|S)
+b	+s	0.60
-b	+s	0.40
+b	-s	0.80
-b	-s	0.20

X	T	C	P(X|T,C)
+x	+t	+c	0.70
-x	+t	+c	0.30
+x	+t	-c	0.70
-x	+t	-c	0.30
+x	-t	+c	0.70
-x	-t	+c	0.30
+x	-t	-c	0.02
-x	-t	-c	0.98

L	B	C	P(L|B,C)
+l	+b	+c	0.60
-l	+b	+c	0.40
+l	+b	-c	0.60
-l	+b	-c	0.40
+l	-b	+c	0.60
-l	-b	+c	0.40
+l	-b	-c	0.01
-l	-b	-c	0.99

Figure 1: Bayes network for the diagnosis

Question 2

If the patient is not smoking and has not recently visited Asia, can you infer with disease?
If the status of nodes A and S is known as $-a$ and $-s$, the probabilities for nodes T, C and B can be derived directly from the conditional probabilities and conditional independance:

- $\mathbb{P}(+t|-a,-s) = \mathbb{P}(+t|-a) = 0.01$
- $\mathbb{P}(+c|-a,-s) = \mathbb{P}(+c|-s) = 0.02$
- $\mathbb{P}(+b|-a,-s) = \mathbb{P}(+b|-s) = 0.80$

Bronchitis is therefore the most likely diagnosis, with 80% probability. There is however a small chance that the patient has a severe disease (cancer or TB)

Question 3

According to the disease inferred in Point 2, the doctor decides to auscultate the patient's lungs with a stethoscope? Why?

Auscultation of the lungs has a low selectivity but high specificity. There are good chances that a sick patient may be undetected (i.e. a negative test is by no mean a guarantee of not being sick). However, the proability of test being positive for a healthy patient is very low. A positive test would thus confirm the diagnosis of either Bronchitis or Cancer.

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

The extended form of the Bayes theorem for a partition $\{A_i\}$ of the sample space is:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum\limits_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)} \tag{1}$$

Equation (1) is applied to the compute the joint probability (B, C) conditioned to L and S from the probability of L conditioned to (B, C, S):

$$\mathbb{P}(B, C|L, S) = \frac{\mathbb{P}(L|B, C, S)\mathbb{P}(B, C|S)}{\sum\limits_{b,c} \mathbb{P}(L|b, c, S)\mathbb{P}(b, c|S)} \tag{2}$$

$$\mathbb{P}(B, C|L, S) = \frac{\mathbb{P}(L|B, C)\mathbb{P}(B|S)\mathbb{P}(C|S)}{\sum\limits_{b,c} \mathbb{P}(L|b, c)\mathbb{P}(b|S)\mathbb{P}(c|S)} \tag{3}$$

Equation (3) is obtained from (2) using the independance of L and S conditional to (B, C), and the independance of B and C continional to S. This is applied to the specific case of a non-smoking patient that has a negative auscultation:

$$\mathbb{P}(B, C| -l, -s) = \frac{\mathbb{P}(-l|B, C)\mathbb{P}(B| -s)\mathbb{P}(C| -s)}{\sum\limits_{b,c} \mathbb{P}(-l|b, c)\mathbb{P}(b| -s)\mathbb{P}(c| -s)} \tag{4}$$

All of the conditional probabilities in (4) are known from the Bayes Network. We can calculate the product $\mathbb{P}(-l|B, C)\mathbb{P}(B| -s)\mathbb{P}(C| -s)$ for each (B, C), and then normalize by the sum to retreive probabilities:

B	C	$\mathbb{P}(-l\|B, C)$	$\mathbb{P}(B\| -s)$	$\mathbb{P}(C\| -s)$	product	$\mathbb{P}(B, C\| -l, -s)$
+b	+c	0.40	0.80	0.02	0.006	1.24%
+b	-c	0.40	0.80	0.98	0.314	60.82%
-b	+c	0.40	0.20	0.02	0.002	0.31%
-b	-c	0.99	0.20	0.98	0.194	37.63%
				Sum	0.516	100.0%

The marginal distributions of B and C, knowing $-l$ and $-s$, are derived by summing the joint probabilities. The probability of Tuberculosis is independant the auscultation outcome, and remains the same :

B	$\mathbb{P}(B\|-l,-s)$
+b	62.06%
-b	37.94%

C	$\mathbb{P}(C\|-l,-s)$
+c	1.55%
-c	98.45%

T	$\mathbb{P}(T\|-a)$
+t	0.01%
-t	99.99%

Since the selectivity of auscultation is low, not much can be inferred from the negative result, and the probability of Bronchitis and Cancer have not changed much by taking into account that additional evidence.

Question 4

The doctor orders an X-Ray. The X-Ray test is positive. What is the new inferred diagnosis?

Inference by enumeration is done for the joint probability (B, C, T):

$$\mathbb{P}(B, C, T|X, L, A, S) = \frac{\mathbb{P}(X, L|B, C, T, A, S)\mathbb{P}(B, C, T|A, S)}{\sum_{b,c,t} \mathbb{P}(X, L|b, c, t, A, S)\mathbb{P}(b, c, t|A, S)} \tag{5}$$

$$\mathbb{P}(B, C, T|X, L, A, S) = \frac{\mathbb{P}(X|C, T)\mathbb{P}(L|B, C)\mathbb{P}(B|S)\mathbb{P}(C|S)\mathbb{P}(T|A)}{\sum_{b,c,t} \mathbb{P}(X|c, t)\mathbb{P}(L|b, c)\mathbb{P}(b|S)\mathbb{P}(c|S)\mathbb{P}(t|A)} \tag{6}$$

In the specific case of our patient with $(-a, -s, -l, +x)$, this gives:

$$\mathbb{P}(B, C, T| + x, -l, -a, -s) = \frac{\mathbb{P}(X| - c, -t)\mathbb{P}(L| - b, -c)\mathbb{P}(B| - s)\mathbb{P}(C| - s)\mathbb{P}(T| - a)}{\sum_{b,c,t} \mathbb{P}(X|c, t)\mathbb{P}(L|b, c)\mathbb{P}(b| - s)\mathbb{P}(c| - s)\mathbb{P}(t| - a)} \tag{7}$$

Individual factors are known from the Bayes network definition, and the product $\mathbb{P}(X| - c, -t)\mathbb{P}(L| - b, -c)\mathbb{P}(B| - s)\mathbb{P}(C| - s)\mathbb{P}(T| - a)$ is computed for all combinations of (B, C, T). Each product is normalized by the sum of products to get a probability.

B	T	C	$\mathbb{P}(-l\|B,C)$	$\mathbb{P}(+x\|T,C)$	$\mathbb{P}(T\|-a)$	$\mathbb{P}(C\|-s)$	$\mathbb{P}(B\|-s)$	product (10^{-3})	$\mathbb{P}(B,C,T\|$ $-l,-s,-a,+x)$
+b	+t	+c	0.40	0.70	0.01	0.02	0.80	0.045	0.23%
+b	+t	-c	0.40	0.70	0.01	0.98	0.80	2.195	11.43%
+b	-t	+c	0.40	0.70	0.99	0.02	0.80	4.435	23.09%
+b	-t	-c	0.40	0.02	0.99	0.98	0.80	6.209	32.33%
-b	+t	+c	0.40	0.70	0.01	0.02	0.20	0.011	0.06%
-b	+t	-c	0.99	0.70	0.01	0.98	0.20	1.358	7.07%
-b	-t	+c	0.40	0.70	0.99	0.02	0.20	1.109	5.77%
-b	-t	-c	0.99	0.02	0.99	0.98	0.20	3.842	20.01%
							Sum	19.205	100.0%

The marginal distributions of B, C and T conditional to $(-l, -s, -a, +x)$ are derived by summing the joint probabilities:

B	$\mathbb{P}(B\| -l, -s, -a, +x)$
+b	67.09%
-b	32.91%

C	$\mathbb{P}(C\| -l, -s, -a, +x)$
+c	29.16%
-c	70.84%

T	$\mathbb{P}(T\| -l, -s, -a, +x)$
+t	18.79%
-t	29.16%

Question 5

Was the X-Ray needed?

Yes! The positive outcome of the X-Ray changed the diagnosis, as there is now a much higher probability that the patient suffers from cancer or tuberculosis. Since that test has a high specificity, much can be inferred from a positive result.