

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

Fundamentals of Artificial Intelligence and Knowledge Representation - Module 3

Bayesian Network for respiratory medicine

Marco Costante - marco.costante@studio.unibo.it,
Alessandra Stramiglio - alessandr.stramiglio@studio.unibo.it

October, 2021

Academic Year 2021/2022

Contents

1	Introduction	2
2	Model	2
3	CPD	3
3.1	Simple CPD	3
3.2	CPD with evidence	4
4	Independencies	4
4.1	Markov Blanket	5
5	Queries	6
5.1	Exact Inference	6
5.1.1	Variable Elimination	6
5.1.2	Map queries	7
5.2	Approximate Inference	7
5.2.1	Rejection sampling	8
5.2.2	Likelihood inference	8
6	Conclusion	9

1 Introduction

In this example we will show a Bayesian network for medical diagnosis of the respiratory system. A Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph. They allow to reason with uncertainty. They are ideal for taking an event that occurred and predicting the likelihood that anyone of the several possible known causes caused the event. In this example, given symptoms, the network can be used to compute the probabilities of the presence of various diseases. The example is taken from a chapter [1] of the book Interactive Collaborative Information Systems [2]. We have chosen this model because Bayesian Networks are useful in diagnosis for medical purpose.

2 Model

The main functionality of the application of the following model is to list the most probable diseases given as inputs the patient symptoms and the background revealing risk factors.

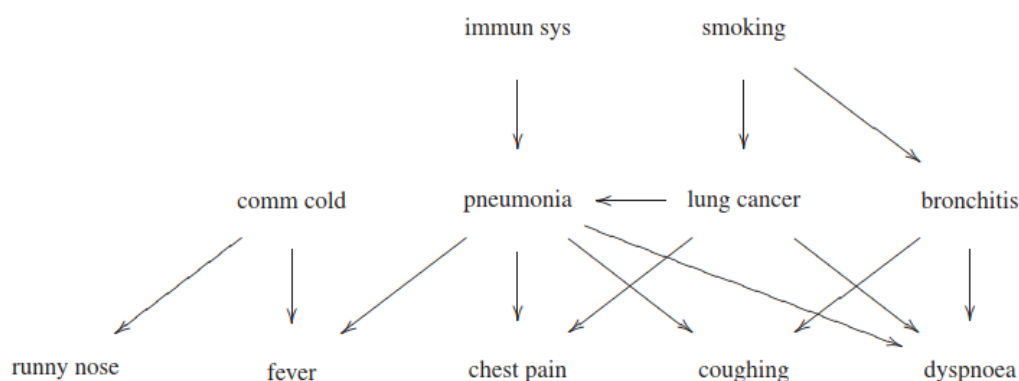
Bayesian networks are a formalism of knowledge representation to express domains that can be characterized by probability distribution. This is a way to capture all the independencies in the domain and hence for compact specification of full joint distributions. A Bayesian network is a directed, acyclic graph with a set of nodes, one per variable, and a conditional distribution for each node given its parents: $P(X_i | Parents(X_i))$. Nodes are connected to each other through edges, which capture independencies.

Now to build a model, we first have to find out which are the variables. In a realistic medical application, one may want to model multi-state variables. For simplicity, however, we take in this example all variables binary (*true/false*). Note that by modeling diseases as separate variables rather than by mutually exclusive states in a single disease variable, the model allows diseases to co-occur.

In this model, we used the following knowledge base. The variables are shown in bold.

- The symptom **dyspnea** (shortness of breath) may be due to the diseases **pneumonia**, **lung cancer** and/or **bronchitis**;
- Patients with **pneumonia**, and/or **bronchitis** often have a very nasty wet coughing;
- **Pneumonia**, and/or **lung cancer** are often accompanied by a heavy **chest pain**;
- **Pneumonia** is often causing a **severe fever**, but this may also be caused by a **common cold**.
- However, a **common cold** is often recognized by a **runny nose**.

- Sometimes, **wet coughing**, **chest pain** and/or **dyspnea** occurs unexplained, or are due to another cause, without any of these diseases being present.
- Sometimes diseases co-occur.
- A **weakened immune-system** (for instance homeless people, or HIV infected) and lung cancer increases the probability of getting an pneumonia;
- **smoking** is a serious risk factor for bronchitis and for lung cancer;



3 CPD

Given two jointly distributed random variables X and Y , the conditional probability distribution of Y given X is the probability distribution of Y when X is known to be a particular value. Conditional Probability Tables (CPT) are an easy way to represent Conditional Probability Distributions (CPD) in form of table.

3.1 Simple CPD

Each variable X in the model is associated with a conditional probability distribution (CPD) that specifies a distribution over the values of X given each possible joint assignment of values to its parents in the model. For a node with no parents, the CPD is conditioned on the empty set of variables. Hence, the CPD turns into a marginal distribution, such as $P(D)$.

Immune system (True)	0.05
Immune system (False)	0.95
Smoking (True)	0.3
Smoking (False)	0.7
Common cold (True)	0.35
Common cold (False)	0.65

Table 1: Simple CPDs

3.2 CPD with evidence

For a node X with parents, we have to consider the conditional probability distribution with the evidence of the value of the parent node Y. So the probability of X to occur is the probability of event X occurring, given that event Y occurs.

Smoking	Smoking (True)	Smoking (False)
Lung cancer (True)	0.1	0.01
Lung cancer (False)	0.9	0.99

Table 2: CPD of having **lung cancer** given **smoking**

Pneumonia	Pneumonia (T)	Pneumonia (T)	Pneumonia(F)	Pneumonia(F)
Bronchitis	Bronchitis (T)	Bronchitis (F)	Bronchitis(T)	Bronchitis(F)
Coughing (True)	0.9	0.9	0.9	0.1
Coughing (False)	0.1	0.1	0.1	0.9

Table 3: CPD of **coughing** given **pneumonia** and **bronchitis**

4 Independencies

Independence can be defined in 3 different ways:

- A is independent of B if the probability of A is independent from B;
- viceversa as before;
- probability of having both A and B is the same of having the product of the probability of A and the probability of B.

In our network there can be made 8602 valid independence assertions, with respect to the all possible given evidence. The main local independencies are:

- $P \models (\text{Immune system} \vdash \text{Bronchitis}, \text{Smoking}, \text{Runny nose}, \text{Common cold}, \text{Lung cancer})$
- $P \models (\text{Smoking} \vdash \text{Immune system}, \text{Runny nose}, \text{Common cold})$
- $P \models (\text{Common cold} \vdash \text{Immune system}, \text{Pneumonia}, \text{Smoking}, \text{Lung cancer}, \text{Coughing}, \text{Chest pain}, \text{Bronchitis}, \text{Dyspnea})$
- $P \models (\text{Pneumonia} \vdash \text{Runny nose}, \text{Common cold}, \text{Bronchitis}, \text{Smoking} \mid \text{Immune system}, \text{Lung cancer})$
- $P \models (\text{Lung cancer} \vdash \text{Immune system}, \text{Runny nose}, \text{Common cold}, \text{Bronchitis} \mid \text{Smoking})$
- $P \models (\text{Bronchitis} \vdash \text{Immune system}, \text{Fever}, \text{Pneumonia}, \text{Chest pain}, \text{Runny nose}, \text{Common cold}, \text{Lung cancer} \mid \text{Smoking})$
- $P \models (\text{Runny nose} \vdash \text{Immune system}, \text{Fever}, \text{Pneumonia}, \text{Smoking}, \text{Lung cancer}, \text{Coughing}, \text{Chest pain}, \text{Bronchitis}, \text{Dyspnea} \mid \text{Common cold})$
- $P \models (\text{Fever} \vdash \text{Immune system}, \text{Smoking}, \text{Lung cancer}, \text{Coughing}, \text{Chest pain}, \text{Runny nose}, \text{Bronchitis}, \text{Dyspnea} \mid \text{Common cold}, \text{Pneumonia})$
- $P \models (\text{Chest pain} \vdash \text{Immune system}, \text{Fever}, \text{Smoking}, \text{Common cold}, \text{Coughing}, \text{Runny nose}, \text{Bronchitis}, \text{Dyspnea} \mid \text{Lung cancer}, \text{Pneumonia})$
- $P \models (\text{Coughing} \vdash \text{Immune system}, \text{Fever}, \text{Smoking}, \text{Common cold}, \text{Lung cancer}, \text{Chest pain}, \text{Runny nose}, \text{Dyspnea} \mid \text{Bronchitis}, \text{Pneumonia})$
- $P \models (\text{Dyspnea} \vdash \text{Immune system}, \text{Coughing}, \text{Fever}, \text{Smoking}, \text{Chest pain}, \text{Runny nose}, \text{Common cold} \mid \text{Lung cancer}, \text{Bronchitis}, \text{Pneumonia})$

4.1 Markov Blanket

The Markov blanket of a given node (in red) consists in the set (in yellow) of:

- Its parents;
- Its children;
- Its children's other parents.

Each node is conditionally independent of all others, given its Markov Blanket.

Once you know this kind of evidence, whatever happens outside does not influence this.

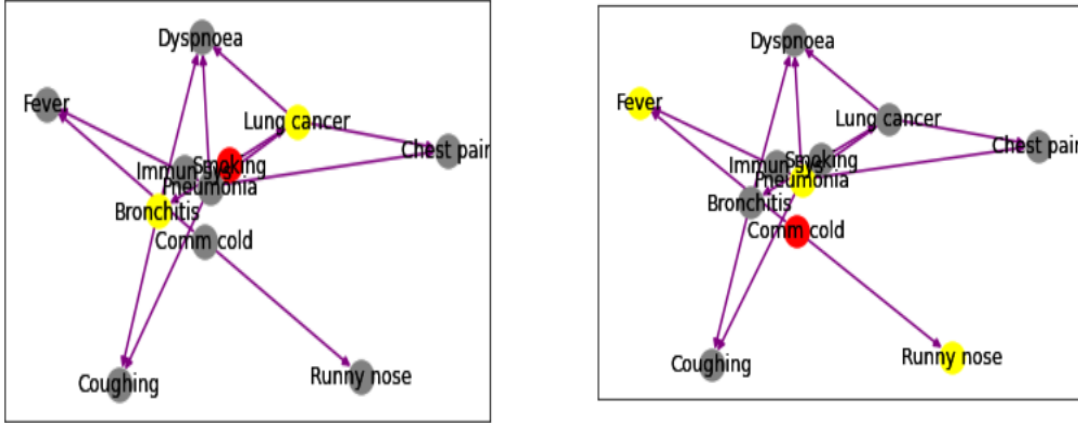


Figure 1: Markov Blanket of *smoking* and *common cold*

5 Queries

Starting from an observed event, it's possible to compute the posteriors probability distribution for a set of query variables, which indeed is the basic task of any probabilistic inference system. In order to do so we need to assign some values to a set of evidences.

5.1 Exact Inference

5.1.1 Variable Elimination

For the exact inference the most common and used method is the one of **variable elimination**, which computes subexpressions only once.

This inference method is called Variable Elimination because it eliminates one by one those variables which are irrelevant for the query. First, the algorithm acts on a set of factors. Each factor f involves a set of variables $\{X_1, X_2, \dots, X_k\}$ and maps instantiations of those variables to real-numbers. The initial set of factors are the network's conditional probability distributions (CPTs). Elimination is driven by an ordering on the variables called an elimination order. During the algorithm, two factor operations are performed many times: factors are multiplied and a variable is summed out of a factor. These factor operations reduce to performing many multiplication and addition operations on real-numbers.

Chest pain and dyspnea can both be caused by lung cancer. However, chest pain for example, can also be caused by pneumonia. The absence of in particular fever makes pneumonia less likely and therefore lung cancer more likely. To a lesser extend this holds for absence of coughing and bronchitis, as we can tell from *Table 5* and *Table 6*.

Pneumonia	$\Phi(\text{Pneumonia})$
Pneumonia (True)	0.4499
Pneumonia (False)	0.5501

Table 4: $P(\text{pneumonia} \mid \text{fever} = \text{True}, \text{runny nose} = \text{False})$

Lung cancer	$\Phi(\text{Lung cancer})$
Lung cancer (True)	0.5686
Lung cancer (False)	0.4314

Table 5: $P(\text{lung cancer} \mid \text{chest pain} = \text{True}, \text{dyspnea} = \text{True}, \text{fever} = \text{False}, \text{coughing} = \text{False})$

Pneumonia	$\Phi(\text{Pneumonia})$
Pneumonia (True)	0.0047
Pneumonia (False)	0.9953

Table 6: $P(\text{Pneumonia} \mid \text{chest pain} = \text{True}, \text{dyspnea} = \text{True}, \text{fever} = \text{False}, \text{coughing} = \text{False})$

5.1.2 Map queries

The following queries are useful for diagnosis purposes and are done with **map queries**. In a map query we have a set of variables like Y and we want to know what is the most probable assignment of this set Y given the evidence. We can take into account this example from the notebook:

- Diagnosis given pneumonia and smoking 'Bronchitis': 'False', 'Comm cold': 'False', 'Lung cancer': 'True'

We see that there is a high probability of lung cancer. The reason is that due to smoking, the prior of disease is increased. More importantly, however, is that weakened immune system is excluded as cause of the pneumonia, so that lung cancer remains as the most likely explanation of the cause of the recurrent pneumonia.

5.2 Approximate Inference

Exact inference for small networks is very efficient and fast, on the contrary approximate inference can take a while. Approximate inference are used in networks with some distributions that are not represented by discrete variables, and so the factor products are not applicable anymore. The idea behind approximate inference is very simple: in stochastic simulations you draw a number of samples from a sampling

distribution, then you compute an approximate posterior probability. Hence for approximate inference we are actually using a frequentist approach rather than bayesian one.

5.2.1 Rejection sampling

Rejection sampling is a general method for producing samples from hard-to-sample distribution given an easy-to-sample distribution. In its simplest form, it can be used to compute conditional probabilities. The algorithm first generates samples from the prior distribution specified by the network. Then, it rejects all those that not match the evidence. Finally, the estimate is obtained by counting how often the evidence occurs in the remaining samples.

5.2.2 Likelihood inference

Likelihood inference avoids the inefficiency of rejection sampling by generating only events that are consistent with the evidence. The algorithm fixes the values for the evidence variables and samples only non-evidence variables. This guarantees that each event generated is consisted with the evidence. Nevertheless, not all events are equal. Before keeping track of the counts in the distribution of the query variable, each event is weighted by the likelihood that the event accords to the evidence. This is measured by the product of the conditional probabilities for each evidence variable, given its parents. Intuitively, events in which the actual evidence appears unlikely should be given less weight.

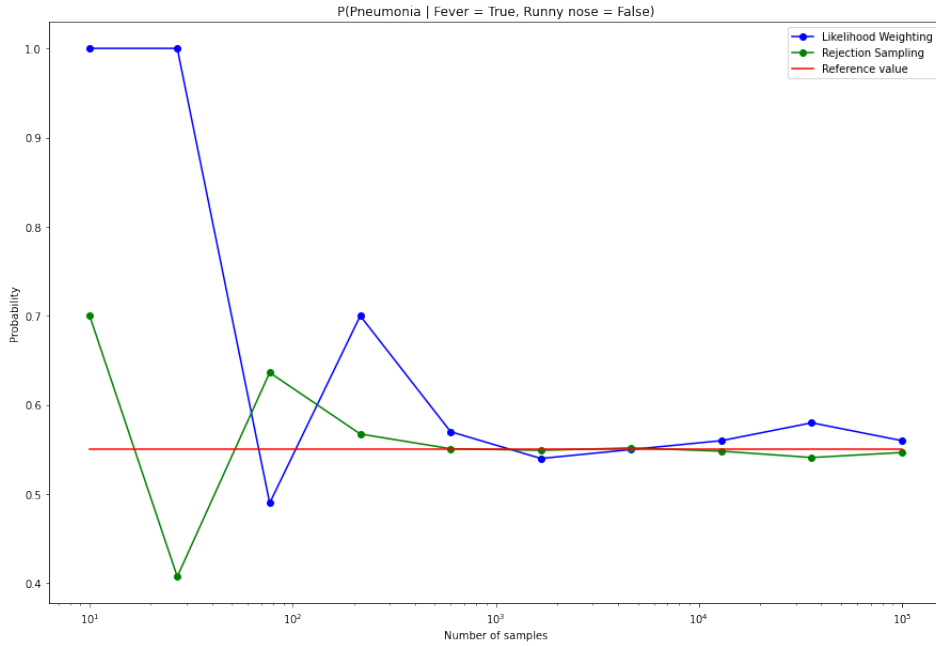


Figure 2: Approximate inference performance comparison

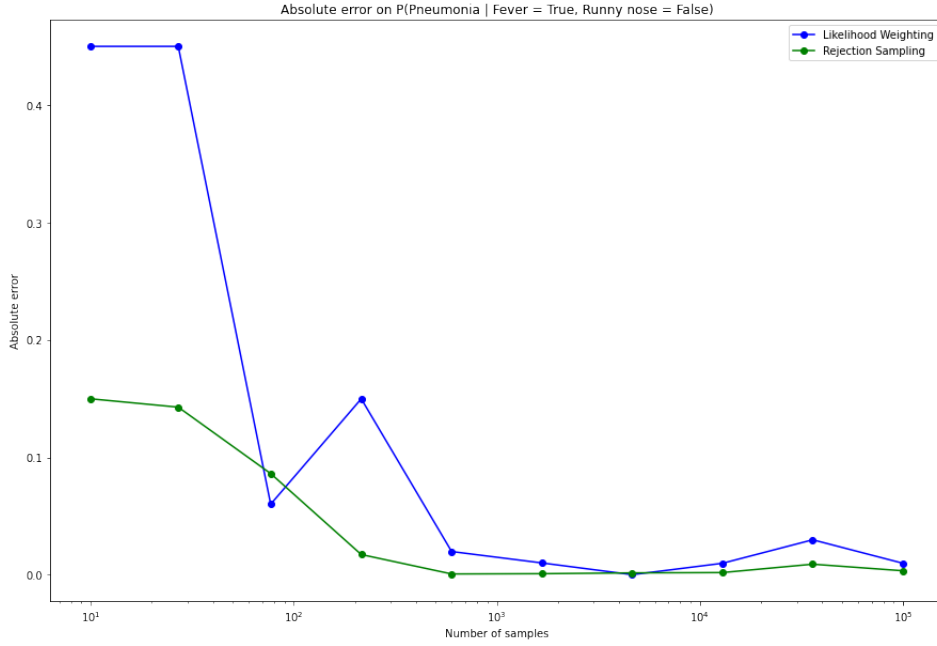


Figure 3: Absolute error of approximate inference

6 Conclusion

As we have said above and as we can see from the graphs, the approximate inference method converges to the reference value after a certain number of samples. Thus we can conclude that in our study case, the approximate method doesn't turn out to be very useful since the developed model is kinda small. In conclusion we can assert that the exact inference method is appropriate.

References

- [1] Bert Kappen Wim Wiegerinck Willem Burgers. *Bayesian Networks for Expert Systems: Theory and Practical Applications*. URL: https://www.researchgate.net/publication/225639547_Bayesian_Networks_for_Expert_Systems_Theory_and_Practical_Applications/.
- [2] Bert Kappen Wim Wiegerinck Willem Burgers. *Interactive Collaborative Information Systems*. URL: https://www.researchgate.net/publication/265540272_Interactive_Collaborative_Information_Systems.