Bayesian Network Mini-Project Report

Andrea Espis
andrea.espis2@studio.unibo.it

1 Introduction

In this project, the domain represented by the Bayesian Network (BN) is an extreme simplification of the cause and effects regarding the Covid-19 virus.

2 Network

2.1 Variables

The variables used to represent the domain are 6 Boolean variables:

- **Contact**: Have/not have been in contact with a person having Covid-19.

- **Mask**: Always/never wearing a mask when not at home.

- **Covid**: Infected/not infected by Covid-19.

- **Fever**: Having/not having fever.

- **Dry_cough**: Having/not having dry cough.

- **Death**: Dying/not dying today.

2.2 Edges

The variables have been ordered following a "cause → effect" reasoning: {Contact, Mask, Covid, Fever, Dry_cough, Death}. The edges representing direct dependencies between the variables are showed in the image below:
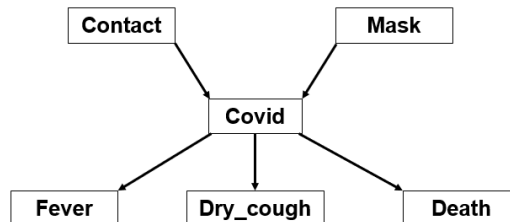
2.3 CPTs

The CPTs have been filled with data from different sources [5] [2] [1] [3] [4] .
I will not discuss here the data gathering since it was not a topic of the course and it was asked a short report. Here are showed the CPTs obtained:

P(Contact)
0.005

P(Mask)
0.65

Covid	P(Fever|Covid)
True	0.88
False	0.015

Covid	P(Dry_cough|Covid)
True	0.67
False	0.25

Contact	Mask	P(Covid|Contact,Mask)
True	True	0.15
True	False	0.85
False	True	0
False	False	0

Covid	P(Death|Covid)
True	0.0018375
False	0.00175

3 Queries

3.1 Exact inference

To do exact inference it has been used the Variable Elimination algorithm, which is more efficient wrt the inference by enumeration.

3.1.1 Causal queries

The causal queries are those in which the evidence variables are the causes, while the query variables are the effects. Some examples of causal queries:

- Chances of having Covid-19, knowing only that you always/never use the mask:
  P(Covid = True |Mask = True) = 0.75%.
  P(Covid = True |Mask = False) = 4.5%.
  In other words, according to this BN, if you never use the mask you have **6** times more chances of getting the virus.

- Chance of dying today, knowing only that you always/never use the mask:
  P(Death = True |Mask = True) =0.1751%.
  P(Death = True |Mask = False) = 0.1754% .
  In other words, according to this BN, if you never use the mask you have **1.0019** times more chances of dying today.

3.1.2 Evidential queries

The evidential queries are those in which the evidence variables are the effects, while the query variables are the effects. Some examples of evidential queries:

- **P(Covid = True |Fever = True)** = Chance of having Covid-19 given that you have the fever = **55.27%**.

- **P(Covid = True |Dry_cough = True)** = Chance of having Covid-19 given that you have dry cough = **5.34%**.

- **P(Covid = True |Fever = True, Dry_cough = True)** = Chance of having Covid-19 given that you have dry cough AND the fever = **76.80%**.

3.2 Approximate inference

It has been done a comparison between the Likelihood Weighting (LW) and Rejection sampling techniques in terms of number of samples needed to converge and absolute error. They have been applied for 4 queries:

- An evidential query:
  P(Covid = True |Fever = True, Dry_cough = True) = 77%.

- A causal query:
  P(Covid = True |Contact = True, Mask = True) = 15% .

- A query with probability close to 0:
  P(Death = True |Mask = True) = 0.1751%.

- A query with many evidences:
  P(Mask = True |Covid = True, Contact = True, Fever = True, Dry_cough = True) = 24%.

The comparison is discussed in the section Results.

4 Methods

The Bayesian Network (BN) discussed has been implemented using the library "pgmpy". The code has been developed by exploiting both the documentation and the notebooks kindly made available by the Professor Paolo Torroni. In particular it has been re-adapted and used the function "run_experiment" to do the comparison between the different sampling techniques.

5 Results

The exact inferences give reasonable probability values, anyway since the BN doesn't faithfully represent the domain, the results must not be trusted of course. About the approximate inferences done using the LW and the Rejection we can summarize the results with these plots:
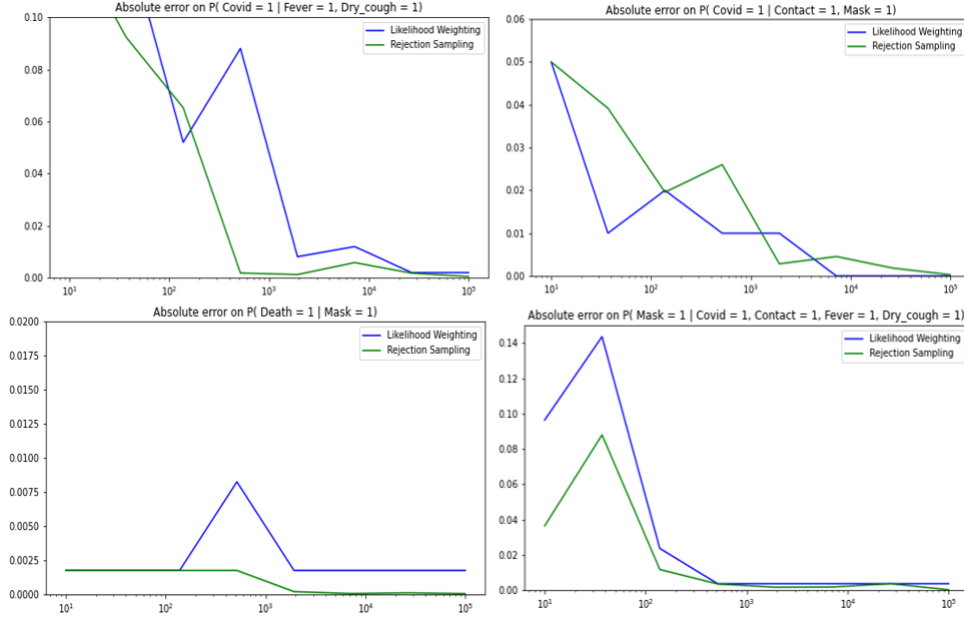
As we expect we can see that for the query P(Death = True |Mask = True) which has probability close to 0, there's a strange behaviour indeed the inference done by LW doesn't converge also with 10^5 samples. We can also notice that LW gets worse if the evidences are many or when the query is evidential. LW doesn't seem to behave better than Rejection sampling, probably due to the fact that there is a little number of nodes in this BN, so the advantage of LW wrt Rejection sampling is not appreciable.

6 Additional analysis

It has been studied the flow of probabilistic influence within the BN with some commands belonging to the pgmpy library. It has been shown an example of blocking the trial composed by {Fever,Covid,Dry_cough} and how the V-structure composed by {Contact, Covid, Mask} can be "enabled". Furthermore it has been shown how to get the markov blanket of a leaf node variable (for which the method "ge_markov_blanket" doesn't work).

References

[1] S. Huang. Covid-19: Why we should all wear masks. https://medium.com/@Cancerwarrior/covid-19-why-we-should-all-wear-masks-there-is-new-scientific-rationale-280e08ceee71.

[2] O. W. in Data. Global deaths. "https://www.weforum.org/agenda/2020/05/how-many-people-die-each-day-covid-19-coronavirus/".

[3] J. Katz. A detailed map of who is wearing masks in the u.s. https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html.

[4] F. Kauffmann. The epidemiology of cough. https://www.hal.inserm.fr/inserm-00744419/document.

[5] W. H. Organization. The symptoms of coronavirus desease [covid-19]. https://www.europepeace.org/coronavirus-outbreak/.