Fundamentals of Artificial Intelligence
and
Knowledge Representation

Covid-19
Bayesian Network

Marini Luca
University of Bologna
Artificial Intelligence Master Degree course

Contents

1 Problem Description

Since 2020 Covid-19 spread all around the World. Many computer scientists, supported by medical experts, have developed online platforms in which it is possible to assess the probability of being positive based on multiple factors.

This project aims to recreate a simple online Covid-19 test that outputs the probability of being positive by leveraging a Bayesian Network. This Network considers users that live in Italy because some assumptions (the number of persons over 80 years old and the number of vaccinated inhabitants) were made considering Italian data.

Please note that the probabilities that this system outputs should **not** be considered as a medical diagnosis. As previously stated, this project aims to simulate an online Covid-19 test. Therefore, it is just an experiment.

The Bayesian Network is composed of eight boolean variables:

- **Persistent_Cough**: this variable models if a Covid patient has had persistent cough.
- **Loss_Taste**: Covid patient had lost taste.
- **Loss_Smell**: Covid patient had lost smell.
- **Fever_Above_38**: Covid patient has had fever above 38°C.
- **Over_Eighty**: the user is over 80 years old.
- **Vaccinated**: the user is vaccinated.
- **Live_With_Other_People**: the user lives with other people in the same house.
- **Positive_To_Covid**: this variable is used to model the positiveness of the user to the virus.

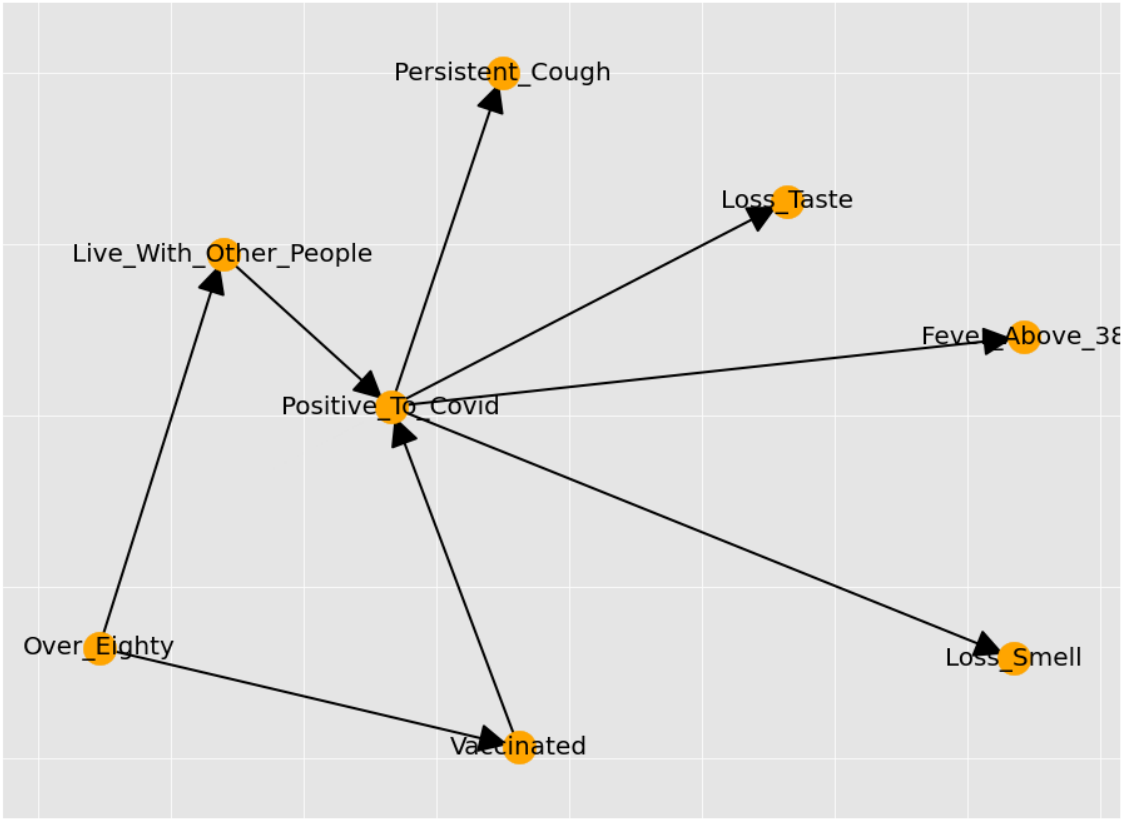

Figure 1: Graphical view of the Bayesian network

1. Being over eighty years old conditions the probability of living with other people in the same house and the probability of being vaccinated. Indeed, in Italy, this age group was given priority to get the vaccine.

2. The two variables Vaccinated and Live_With_Other_People condition the probabilities of Positive_To_Covid.

3. The positiveness to Covid-19 conditions the four symptoms Persistent_Cough, Loss_Taste, Fever_Above_38, and Loss_Smell because it is possible to know the symptoms' probabilities only after collecting data of positive patients.

2 Conditional Probability Tables (CPT)

In this section are reported the Conditional Probability Tables for each node in the Bayesian Network.

Please note that most of the following data is not accurate, while some statistics are taken from the *Italian National Institute of Statistics* (ISTAT).

2.1 CPT - Over_Eighty

Over_Eighty(No)	0.93
Over_Eighty(Yes)	0.07

In March 2021, 4.442.048 people in Italy have more than 80 years old [3]. The total number of Italian inhabitants in 2021 is 59.641.488, as visible on the ISTAT website [2].

2.2 CPT - Vaccinated

Over_Eighty	Over_Eighty(No)	Over_Eighty(Yes)
Vaccinated(No)	0.92	0.57
Vaccinated(Yes)	0.08	0.43

The percentage of over 80 years old Italian people that are vaccinated is 42.6% (1.893.746 out of 4.442.048), while just 8.35% of the whole Italian population has been vaccinated [3].

2.3 CPT - Live_With_Other_People

Over_Eighty	Over_Eighty(No)	Over_Eighty(Yes)
Live_With_Other_People(No)	0.35	0.55
Live_With_Other_People(Yes)	0.65	0.45

In Italy, 44.5% of the elderly do not live alone [1].

2.4 CPT - Positive_To_Covid

Live_With_Other_People	Live_W_Other_People(No)	Live_W_Other_People(No)
Vaccinated	Vaccinated(No)	Vaccinated(Yes)
Positive_To_Covid(No)	0.88	0.98
Positive_To_Covid(Yes)	0.12	0.02

Live_With_Other_People	Live_W_Other_People(Yes)	Live_W_Other_People(Yes)
Vaccinated	Vaccinated(No)	Vaccinated(Yes)
Positive_To_Covid(No)	0.82	0.94
Positive_To_Covid(Yes)	0.18	0.06

2.5 CPT - Loss_Taste

Positive_To_Covid	Positive_To_Covid(No)	Positive_To_Covid(Yes)
Loss_Taste(No)	0.9	0.43
Loss_Taste(Yes)	0.1	0.57

2.6 CPT - Loss_Smell

Positive_To_Covid	Positive_To_Covid(No)	Positive_To_Covid(Yes)
Loss_Smell(No)	0.85	0.46
Loss_Smell(Yes)	0.15	0.54

2.7 CPT - Fever_Above_38

Positive_To_Covid	Positive_To_Covid(No)	Positive_To_Covid(Yes)
Fever_Above_38(No)	0.82	0.3
Fever_Above_38(Yes)	0.18	0.7

2.8 CPT - Persistent_Cough

Positive_To_Covid	Positive_To_Covid(No)	Positive_To_Covid(Yes)
Persistent_Cough(No)	0.75	0.2
Persistent_Cough(Yes)	0.25	0.8

3 Independence

In this section are shown the local independence of the two variables Vaccinated and Positive_To_Covid.

```
(Vaccinated _|_ Live_With_Other_People | Over_Eighty)
```

```
(Positive_To_Covid _|_ Over_Eighty | Live_With_Other_People, Vaccinated)
```

4 Inference

It is possible to infer the probabilities of specific nodes of the Bayesian Network given other variables' values. Moreover, there are two main Inference methods:

- Exact Inference;
- Approximate Inference.

4.1 Exact Inference

The method used to perform Exact Inference is Variable Elimination.

4.1.1 Probability of being positive to Covid-19

The following table shows the probability of being positive to Covid-19 without any given assumption.

```
P(Positive_To_Covid)
```

```
+------------------------+--------------------------+
| Positive_To_Covid      |   phi(Positive_To_Covid) |
+========================+==========================+
| Positive_To_Covid(No)  |                   0.8535 |
+------------------------+--------------------------+
| Positive_To_Covid(Yes) |                   0.1465 |
+------------------------+--------------------------+
```

The probability of being positive without any given assumption (14.65%) is lower than being negative (85.35%).

4.1.2 Probability of being positive to Covid-19 having loss the sense of smell

```
P( Positive_To_Covid | Loss_Smell = Yes )
```

```
+------------------------+--------------------------+
| Positive_To_Covid      |   phi(Positive_To_Covid) |
+========================+==========================+
| Positive_To_Covid(No)  |                   0.2212 |
+------------------------+--------------------------+
| Positive_To_Covid(Yes) |                   0.7788 |
+------------------------+--------------------------+
```

The probability of being positive to Covid-19, given the loss of smell, is 77.88%.

4.1.3 Probability of being positive to Covid-19 being vaccinated

```
P( Positive_To_Covid | Vaccinated = Yes )
```

```
+-----------------------+--------------------------+
| Positive_To_Covid     |   phi(Positive_To_Covid) |
+=======================+==========================+
| Positive_To_Covid(No) |                   0.9563 |
+-----------------------+--------------------------+
| Positive_To_Covid(Yes) |                  0.0437 |
+-----------------------+--------------------------+
```

If one person is vaccinated, then the probability of being infected with the virus falls to 4%.

4.1.4 Maximum A-posteriori Probability

In this subsection, the Maximum A-posteriori Probability (MAP) of the variable Positive_To_Covid is computed.

It can be noticed that if someone has the virus, then the most probable values of the variables Over_Eighty, Live_With_Other_People, and Vaccinated are respectively No, Yes, and No. Hence, it means that it is more probable that a person is positive to Covid-19 if he/she:

- is younger than 80 years old rather than older;
- lives with other people rather than living alone;
- is not vaccinated.

```
MAP of Positive_To_Covid = 'Yes'
```

```
{'Over_Eighty': 'No', 'Live_With_Other_People': 'Yes', 'Vaccinated': 'No'}
```

4.2 Approximate Inference

In this section, Approximate Inference is applied to compute the probability of the variable Vaccinated, given that the user is older than 80 years old and that has not the virus.

The same probability is first calculated with Exact Inference (Variable Elimination) to visualize how well the Approximate methods perform compared to the Exact Inference methods.

```
Exact Inference:

P(Vaccinated | Over_Eighty = "Yes", Positive_To_Covid = "No")

+-----------------+-------------------+
| Vaccinated      |   phi(Vaccinated) |
+=================+===================+
| Vaccinated(No)  |            0.5403 |
+-----------------+-------------------+
| Vaccinated(Yes) |            0.4597 |
+-----------------+-------------------+
```

The 45.97% of over eighty Italian people that are not positive to Covid-19 are vaccinated.

Then, the probability P(Vaccinated | Over_Eighty = "Yes", Positive_To_Covid = "No") is estimated by using two different Approximate Inference methods, which are:

- Weighted Likelihood;
- and Rejection Sampling.

In Figure 2 it can be seen that they are tested with an increasing number of samples.

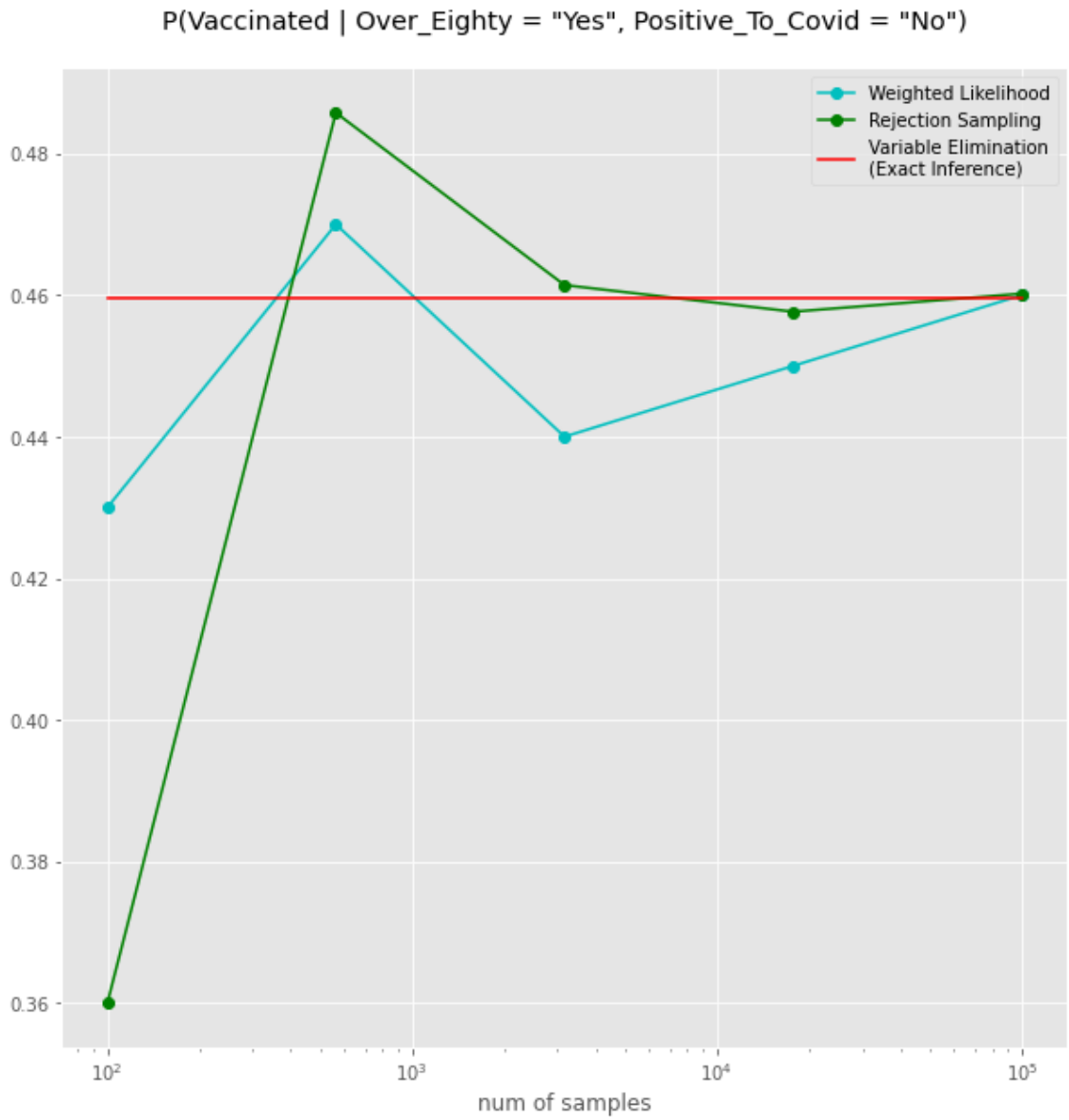

Figure 2: Test of three different Approximate Inference methods

It is visible that, in five steps, the two Approximate Inference methods bring great results close to the Variable Elimination one.

The most accurate (the one with the minor error) is the Weighted Likelihood method, which approximates the probability P(Vaccinated | Over_Eighty = "Yes", Positive_To_Covid = "No") to 46%.

References

[1] *Aspetti di vita degli over 75*. URL: https://www.istat.it/it/files/2020/04/statisticatoday_ANZIANI.pdf.

[2] *Popolazione Italiana residente al 1° gennaio*. URL: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1.

[3] *Vaccini in tempo reale*. URL: https://lab24.ilsole24ore.com/numeri-vaccini-italia-mondo.