Bayesian Theorem
Assignment 2

Prof.	Panos Trahanias	trahania@csd.uoc.gr
T.A.	Emmanouil Sylligardos	sylligardos@csd.uoc.gr
T.A.	Despina - Ekaterini Argiropoulos	despargy@csd.uoc.gr
Deadline	**Wednesday 23:59, 22nd of March 2023**	

Note: This is a personal assignment and should be pursued individually and without use of any computerised AI facilities. Note that this will be automatically checked for every delivered assignment. The assignment should be implemented entirely in Google Colaboratory following the delivered instructions below.

Part A: (30/100)

Data - Part A: the total disease cases in 1996 were approximately 290,000 worldwide,
the world population in 1996 was estimated to be $6,8 \cdot 10^9$ people,
sensitivity of the test = 0.99%, specificity of the test = 0.96% .

Equation - Part A: For Part A, you should use the equation 1, where Test positive (T^+), Test negative (T^-), Disease present (D^+), Disease absent (D^-).

$$P(D^+|T^+) = \frac{P(T^+|D^+) \cdot P(D^+)}{P(T^+)} \tag{1}$$

Question 1:

1. Implement a function that takes as input the sensitivity, the specificity of a test and the prior probability of having a disease and returns the posterior probability of having that disease given that a person was tested positive.
2. Compute the probability of that person actually having the disease given that it tested positive.
3. Save and print the posterior probability of that person having the disease given that another test has been done and it turned to be positive.
4. Compute the probability of that person actually having the disease (use the first posterior you computed), for each one of the following cases: when the total cases in 1997 were (a) 2 million, (b) 7.5 million, (c) 1 billion.
5. Briefly explain how/why the posterior probability of actually having the disease given that you tested positive, is affected by the prior.

Part B: (70/100)

In a classroom, there are 23 children. The teacher decides to find out about their birthday. Notice that the teacher cares only about the day and the month of their birthday (365 days in total), not the year. Each child's birthday is equally likely to be any of the 365 days, independently.

Question 1:

1. Compute the probability of 2 children having different birthdays.
2. Compute the number of all the possible combinations for a pair of 2 children in that class. In other words, how many comparisons (combinations) should be made to compare 23 birthdays against each other?

3. Implement a function to compute the probability of all children having different birthdays. The function should have as input the total number (m) of children in the class (m = 23 in our case) and return the probability. Again each child is equally likely to be born on any of the 365 days regardless of the birthdays of the others.
4. How many children should there be in the class to be mathematically certain that there are at least 2 children with the same birthday. Explain your answer shortly.

Question 2: Bayesian update

The teacher now challenged the children to find her birthday. Each one of the children would get up from their desk, go to the teacher, silently tell her its guess (without the other children listening) and then sit back down to its desk. The teacher would only right on the board whether the guess of the child 'landed' left or right from her birthday.

- A year has 365 days
- Initially, the birthday of the teacher are equally likely to be any day of the year.
- A guess (e.g. 12 of March) is said to land left if the day is earlier than the teacher birthday, and right otherwise.
- In total, there were 23 guesses
- On the board there was written the following:
  L L L R L R R L R R R L L L R L L R R R L L L

You are asked to:

1. Implement a function to compute the probability of the teacher's birthday to be on a day x, given the observations $\{L, N\}$, $P(x|L, N)$. Follow the steps bellow:
   (a) Compute the probability of a single guess being left
   (b) Compute the probability of 5 guesses in a row being left.
   (c) Compute the probability of 4 guesses out of 10 being left.
   (d) Having answered all the previous steps you should now be able to complete the initial question 2.1.
2. Plot the probability distribution for the teacher's birthday after having all the evidence.
3. What's the most probable date of the year for the teacher's birthday.
4. (**10% Bonus**) We would like to visualize how our belief about the teacher's birthday changes with each new observation. Create an animated plot of 23 frames, one frame per guess, of the probability distribution over the days of the year. This plot should show how the probability distribution changes after each new evidence.
   Hints:
   - You should use the Bayes theorem to update your beliefs at each step.
   - Do not use the evidence of the Bayes theorem, that is to say, use only the prior and the likelihood to compute the posterior. The evidence would only scale the shape of the result.

Deliverable

This assignment should be implemented entirely in Google Colaboratory. Google's notebook allows you to combine executable Python scripts with rich text in a single document. Your deliverable should be a single .ipynb file along with its corresponding .py file (both can be easily exported from Google Colaboratory). Every single question should be implemented in a single code block. Code blocks should be clearly and shortly explained (you may use the text boxes for that goal). Use **only** library functions for matrix operations and plots.