Bayesian Theorem
Assignment 2

Prof.	Panos Trahanias	trahania@csd.uoc.gr
T.A.	Myrto Vilia	mvillia@ics.forth.gr
T.A.	Despina - Ekaterini Argiropoulos	despargy@csd.uoc.gr
T.A.	Michalis Savorianakis	csdp1387@csd.uoc.gr
Deadline	**Tuesday 23:59, 1st of April 2025**	

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 and for cases of relevant detections the whole assignment will be dropped. The assignment should be implemented entirely in Google Colaboratory following the delivered instructions below. **Comments regarding the submitted code are mandatory.**

Part A: Medical test (30/100)

Data – Part A: A new medical test has been developed to detect a rare disease. The disease affects approximately 1 in 10,000 people in the general population. The sensitivity of the test is = 78%, specificity of the test = 89% .

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^-).

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

Questions:

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. Calculate the probability that that person actually has the disease.
3. Save and print the posterior probability of that person having the disease, given that another test has been done and it turned out 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 are (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: Startup Company (70/100)

A start-up company just received huge funding, and its boss is so happy that he decided to give each of them a day off in a time window of 40 days. The company's employees are 20 in number ($E_1, E_2, \ldots, E_{20}$) and are all in a conference room ready to select their one day off. Each has to select a number between 1 and 40, independently, without knowing which number the other employee selects.

Questions : Probabilities (20/100)

1. Compute the probability that one employer will pick the number $n \in [1, 40]$ (e.g. $n = 5$), a specific number.
2. Compute the probability of 2 employers picking the same number.

Released date: Tuesday, 18th of March 2025
Deadline: Tuesday 23:59, 1st of April 2025

3. Compute the probability of 2 employees picking a different number.
4. Compute the probability of 3 employees picking a different number.
5. Implement a function to compute the probability that all employees will select different numbers. The function should have as input the total number (m) of people in the room (m = 20 in our case) and return the probability. Again, each employ is equally likely to select any number of the 40, regardless of the choice of the others.
6. Using the previous function, plot the probability of all employees selecting different numbers as a function of the number of employees. The x-axis should represent the number of employees. The y-axis should represent the probability of all employees having unique selections. Explain the results.

Each employee has selected the day-offs and their are all happy. Now, the boss of the above startup company, which has a really nice sense of humor, decides to select a number, as he was one of the employees. If an employee guess correctly this number that his boss picked, the lucky employee takes a second day-off in a row! If the employee's answer is wrong, then the boss says either smaller ('S') or greater ('G'), depending on the number he has in his mind is smaller or greater to the one that one employee said. Each employee is able to listen what the previous employee guessed, in order to find their boss selected number.

Each employee guesses independently (but in public), one at a time, and the boss will indicate if their guess is smaller, greater, or correct. After each guess, the probability distribution of the boss's number is updated using Bayesian updating.

- Number of employees: 20 employees.
- Available numbers: The numbers range from 1 to 40.
- Boss's number: The boss's number is hidden, and all employees are guessing what it might be.
- Feedback: The boss gives one of the following responses after each employee guesses:
  - Smaller: If the boss's number is smaller than the employee's guess.
  - Greater: If the boss's number is greater than the employee's guess.
  - Correct: If the employee guesses the correct number.
- Initial Belief: Each employee starts with the belief that the boss's number is equally likely to be any number between 1 and 40 (uniform prior).
- Observations: After each employee's guess, the boss gives feedback about whether the guess was too high or too low (or correct).
- Boss answers: S S S G S G G S G G S S G S S G G S S S

Questions: Implementing Bayesian update (50/100)

1. Implement a function to compute the probability distribution for the boss's number being a specific value x, given a sequence of observed guesses and feedback (after each employee's guess). Follow the steps below:
   (a) Compute the probability of a single guess being "smaller" and "greater", given a guess x, where x is an integer $x \in [1, 40]$.
   $P(S|x) = ?, \quad P(G|x) = ?$. Write a simple function for each case, "smaller" and "greater".
   (b) Compute the probability that the first 5 guesses **all** resulted in "smaller" feedback.
   $P(S = 5, N = 5|x) = ?$. Write a simple function.
   (c) Compute the probability of 4 guesses out of 10 resulted in "smaller" feedback. (Hint: "from math import comb")
   $P(L = 4, N = 10|x) = ?$. Write a simple function.
   (d) Using the above steps, compute the posterior probability distribution $P(x|observations)$ for the boss's chosen day-off number.
2. Plot the probability distribution of the boss's chosen number after all evidence (guesses and feedback) has been observed. Use "Boss answers".
3. Determine the most probable day-off number for the boss.

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.