Parameter Estimation
Assignment 2

Instructor	Iason Oikonomidis	oikonom@csd.uoc.gr
T.A.	Giorgos Karvounas	gkarv@csd.uoc.gr
Deadline	**Thursday 23:59, 28th of March 2024**	

Abstract

In this assignment the task is to familiarize with Parameter Estimation. The goal of this assignment is to apply the methods we've discussed for parametric density estimation, MLE and Bayesian in real problems.

Note: This is a personal assignment and should be pursued individually and without use of any computerized AI facilities. Note that there will be a face-to-face examination after the delivery. The assignment should be implemented entirely in Google Colaboratory following the deliverable instructions below. *Use library implementations only when explicitly stated*.

Question A: Poisson MLE derivation and application (45/100)

The Poisson distribution models discrete events that happen independently at a given rate λ per time unit. It is parameterized by the rate λ, with PMF: $P(k) = \exp(-\lambda)\ \lambda^k / \kappa!$ where k represents the observed number of events. You may find useful the ability to embed latex code in colab notebooks, though this is not strictly necessary. See an example.

1. Start by writing and expanding the likelihood function for a sample x_1, x_2, …, x_n drawn independently from a Poisson distribution. Briefly comment why and under which assumptions it can take the form you've reached. Afterwards, derive the log-likelihood function and describe the resulting simplifications.
2. Compute the gradient of the log-likelihood function. Since there is a single parameter, this amounts to computing the derivative wrt λ. Set it equal to 0, and find the maximum likelihood estimator λ_{MLE}. Briefly explain all steps.
3. Write a Python function that takes an array of data points from a Poisson-distributed variable, and returns this MLE. Pick a rate λ, use `numpy.random.poisson` as shown below to draw n samples, and use your implemented function to estimate λ. Experiment with different values for λ and n, and describe your results.

```
import numpy as np
samples = np.random.poisson(lam=λ, size=n)
```

Question B: Birthday Problems and Bayes' Rule (65/100)

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 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, and independently of the others. (*just for "fun": how do these assumptions break in the real world?*)

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 so that it would be mathematically certain that two of them share a birthday? Briefly justify your answer, comparing and contrasting with the previous results.

Question 2:

The teacher then challenged the children to find her birthday. The first step was each of them to pick a date in secret. Then, in turn, each child would get up from their desk, go to the teacher, silently tell their guess (without the other children listening) and then sit back down at their desk. The teacher would only write on the board whether the guess of the child 'landed' left or right from her birthday (let's ignore the case where the guess is correct).

- A year has 365 days
- Initially, the birthday of the teacher is 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's birthday, and right otherwise.

In total, there were 23 guesses, and in the end, 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?

Deliverable

This assignment should be implemented entirely in Google Colaboratory. Your deliverable should be a single .ipynb file (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). **Only use library functions for matrix operations and plots, unless explicitly stated.**

Release date: Wednesday, 15th of March 2024
Deadline: Thursday 23:59, 28th of March 2024