Summer-2018

Admission Test for the Data Incubator Reply

Instructions:

- You are allowed 3 hours (180 minutes) for reading, solving, and sending back this exam
- This exam is individual and you are not allowed to solve it with the assistance of another person
- You may use calculator, computer, notes, books, internet, or other printed/online non-personal assistance on this exam
- Please begin every page with your name and the exercise number
- To receive full credit, you must show your work and carefully justify your answers. The correct answer without any work will receive little or no credit
- If you need it, you can use more than one page per exercise. Do not answer two or more exercises in the same page. You may either print the test or use white sheets to write your answers.
- Please write neatly. Illegible answers will be assumed to be incorrect.
- For the programming exercises, we suggest you to use Python. However, you may use any other language or pseudocode to write your solution.
- **Avoid importing additional libraries**
- For every programming exercise, look for the optimal solution in terms of time complexity and memory usage. In case one of them is compromised by the other, select one and explain for which cases the other aspect might have problems.
- Do your best!

(2 points)

1.- A program reads and processes text files every day. The number of files per day is not known in advance, but it can be modeled with a Poisson distribution and rate 5 files/day. When the number of files in a day is larger or equal than 10, then the program fails to read the files due to overload. What is the probability that the program fails at least twice during the next week?

(2 points)

2.- One of the Reply companies uses a keypad (with the numbers 1 to 9) to unlock their office door. They used to have a 4-digit PIN, without repetition of digits. Now, management has decided that this should be changed to a 6-digit PIN, without repetitions, in order to increase security. To help the employees remember the PIN, they decide that the new 6-digit PIN should always contain one of the left-to-right rows or top-to-bottom columns on the keypad, i.e. one of the sequences 123, 456, 789, 147, 258, or 369.

1	2	3
4	5	6
7	8	9

Calculate the number of variations for both the old and the new PIN system. Is the new rule actually more secure?

(1+2 points)

3.- You have a damaged clock that only shows full hours $h \in \{0, 1, \ldots, 11\}$. Every hour the clock advances by 1 with a probability $f \in [0, 1]$.

a) What is the probability $P(h=H \mid t=T)$ that after $T \in \{0, 1, \ldots, 11\}$ hours the clock shows the time $H \in \{0, 1, \ldots, 11\}$?

b) After setting the clock to 0 you go to sleep. When you wake up, the clock shows h=6. You know from experience that you always sleep either exactly 7 or exactly 8 hours (with equal probability). For f=0.6, what is the probability that you slept 7 hours vs. the probability that you slept 8 hours?

(1+1 points)

4.- A milk producer buys a machine that fills 1-liter bottles. The machine is not perfect and sometimes it fills a bit more and sometimes a bit less than 1 liter. The manual says that the standard deviation of the amount is 0.1 liter. The producer decides to test the machine to check if the mean value is 1 liter.

The results of the test with 10 bottles are (in liters): 0.90, 1.10, 0.98, 1.13, 0.95, 1.25, 1.17, 1.25, 0.96, 1.31 (the sample mean is 1.1 liters). If the amount of milk that the machine serves, X, is modeled with a normal distribution with unknown mean μ and standard deviation $\sigma = 0.1$, answer the following questions:

a) Obtain a 95% confidence interval of the mean amount of milk, μ, using the sample mean as your estimator
b) Perform a hypothesis test with 95% confidence and conclude if you should reject or not the null hypothesis $\mathcal{H}_0$: $\mu = 1$ liter. What is the p-value of the test?

Justify any assumption that you had to make to solve these questions.

(2 points)

5.- Write a function that takes as an input two sorted lists (both might be in increasing or decreasing order, or one in increasing and the other one in decreasing) and outputs a list that is their union in decreasing order. DO NOT use any sorting or ordering function.

Example:

INPUT1>> [“aba“, “ba“, “cam“, “pin“]

INPUT2>> [“vat”, “dui”, “bin”]

OUTPUT>> [“vat”, “pin”, “dui”, “cam”, “bin”, “ba”, “aba”]

(2 points)

6.- If we list all the positive integers below $y = 10$ that are multiples of $x_1 = 3$ or $x_2 = 5$, we get 3, 5, 6 and 9. The sum of these multiples is 23. Write a python code that finds the sum of all the multiples of x_1 or x_2 below y.

Example1:

INPUT >> x1=2, x2=7, y=100

OUPUT >> 2793

Example2:

INPUT >> x1=3, x2=5, y=1000

OUPUT >> 233168

(3 points)

7.- You are given a list of filenames, and the number of users U > 2. The list of filenames will always contain a multiple of U*(U-1)/2 elements. Assign the files to the users in the following way:

- Each file should be assigned to exactly two users.
- Each user should share the same number of files with each other user.

Return one of the possible assignments of files to all users.

Bonus question: How many solutions (unique assignments) are there?

Example:

filenames = ['file0.txt', 'file1.txt', 'file2.txt', 'file3.txt', 'file4.txt', 'file5.txt']

U = 3

OUTPUT:

[['file0.txt', 'file1.txt', 'file4.txt', 'file5.txt'],

['file0.txt', 'file1.txt', 'file2.txt', 'file3.txt'],

['file2.txt', 'file3.txt', 'file4.txt', 'file5.txt']]