Tasks: Fundamentals of Data Analysis

ian.mcloughlin@atu.ie

Last updated: 30 August 2023

1. The Collatz conjecture[1] is a famous unsolved problem in mathematics. The problem is to prove that if you start with any positive integer x and repeatedly apply the function $f(x)$ below, you always get stuck in the repeating sequence $1, 4, 2, 1, 4, 2, \ldots$

$$f(x) = \begin{cases} x \div 2 & \text{if } x \text{ is even} \\ 3x + 1 & \text{otherwise} \end{cases}$$

   For example, starting with the value 10, which is an even number, we divide it by 2 to get 5. Then 5 is an odd number so, we multiply by 3 and add 1 to get 16. Then we repeatedly divide by 2 to get 8, 4, 2, 1. Once we are at 1, we go back to 4 and get stuck in the repeating sequence 4, 2, 1 as we suspected.

   Your task is to verify, using Python, that the conjecture is true for the first 10,000 positive integers.

2. Give an overview of the famous penguins data set,[2] explaining the types of variables it contains. Suggest the types of variables that should be used to model them in Python, explaining your rationale.

3. For each of the variables in the penguins data set,[3] suggest what probability distribution from the numpy random distributions list is the most appropriate to model the variable.[4]

4. Suppose you are flipping two coins, each with a probability p of giving heads. Plot the entropy of the total number of heads versus p.

5. Create an appropriate individual plot for each of the variables in the penguin data set.[5]

[1] *The Simple Math Problem We Still Can't Solve | Quanta Magazine*. Sept. 22, 2020. URL: https://www.quantamagazine.org/why-mathematicians-still-cant-solve-the-collatz-conjecture-20200922/ (visited on 08/18/2023).

[2] *mwaskom/seaborn-data: Data repository for seaborn examples*. Aug. 30, 2023. URL: https://github.com/mwaskom/seaborn-data/blob/master/penguins.csv (visited on 08/30/2023).

[3] *mwaskom/seaborn-data: Data repository for seaborn examples*.

[4] *Random Generator — NumPy v1.25 Manual*. June 17, 2023. URL: https://numpy.org/doc/stable/reference/random/generator.html#distributions (visited on 08/30/2023).

[5] *mwaskom/seaborn-data: Data repository for seaborn examples*.