SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

ANL252
Python for Data Analytics

Tutor-Marked Assignment

July 2021 Presentation

TUTOR-MARKED ASSIGNMENT (TMA)

This assignment is worth 18% of the final mark for ANL252 Python for Data Analytics.

The cut-off date for this assignment is **8 August 2021, 2355hrs**.

Up to 25 marks of penalties will be imposed for inappropriate or poor paraphrasing. For serious cases, they will be investigated by the examination department. More information on effective paraphrasing strategies can be found on https://academicguides.waldenu.edu/writingcenter/evidence/paraphrase/effective.

Note to Students:

You are to include the following particulars in your submission: Course Code, Title of the TMA, SUSS PI No., Your Name, and Submission Date.

Question

The probability density function (pdf) of the normal distribution is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

where x is the observed value of a random variable X, π is the value of the constant pi ($\approx$ 3.14159), μ is the mean and σ^2 is the variance of the distribution, and exp() is the exponential function.

As an analyst and Python programming specialist, you are asked to construct a program to compute the probability of a normally distributed random variable given the distribution mean and variance.

Note: Include your Python program code in the answers and show them in the "Consolas" or "Courier New" fonts (size 12). Make a screenshot of the program output if required.

(a) Prepare the programming of the function by importing the "math" package into your program which is needed for the square root and the exponential functions. You can also find a much more exact value for π in the math package. Refer to the official website of the math package https://docs.python.org/3/library/math.html for more details of the functions.

(1 mark)

(b) Employ a Python program to ask the user to enter the mean and variance of the distribution. Tell the user that the mean can be any value between minus infinity ($-\infty$) and plus infinity ($+\infty$), but the variance must be a value larger than 0. Integrate certain control mechanism to ensure that the variance condition is fulfilled, and the input is numeric. If the user press ENTER without providing any values, the program will automatically set μ to 0 and σ^2 to 1.

(15 marks)

(c) Design an input screen for the user to enter the value of X. Tell the user that the value can be any value between minus infinity (–∞) and plus infinity (+∞). Integrate certain control mechanism to ensure that the input is numeric.

(4 marks)

(d) Construct a user-defined function using the formula of the probability density function given above to compute the corresponding probability density $f_X(x)$ based on the user inputs in (b) and (c).

(15 marks)

(e) Use formatted printing to display the result of (d) to the user.

(10 marks)

(f) The probability of a normally distributed random variable X smaller equal to a value k can be determined by the cumulative distribution function (cdf) which can be approximated by the weighted sum of the density function of X from –∞ to k:

$$P(X \leq k) \approx \alpha[f_X(a) + \cdots + f_X(k - 2\alpha) + f_X(k - \alpha) + f_X(k)],$$

where α is the range of each step and a is a number close to –∞. For example, if we choose α = 0.01, a = -100 and k = 0, then the summation is

$$P(X \leq 0) \approx 0.01[f_X(-100) + \cdots + f_X(-0.02) + f_X(-0.01) + f_X(0)]$$

The accuracy of the approximation increases with the decrease of the step range α. That means, the smaller the step width is, the more terms we can put between a and k, and the more accurate is the probability.

Design a program to compute $P(X \leq k)$ where k is the value that the user entered in (c). Use the user input in (b) as the values for the parameters μ and σ^2 and display the result to the user by formatted printing. Furthermore, check the results for k = 0, 1.64, and 1.96 with a mean of 0 and variance of 1 as well.

(25 marks)

(g) Explain your program for the calculation of $P(X \leq k)$ in (f) in less than 200 words.

(15 marks)

(h) Create a dictionary to store the probabilities of the normal distribution for x = {-5, -4.9, -4.8, …, 0, 0.1, 0.2, …, 4.8, 4.9, 5} where x are the keys, and the corresponding probabilities are the values of the dictionary. The distribution mean here is 0 and variance is 1. Print all the probabilities (with the corresponding x) of those x's between -2 and 2 with a step width of 0.5 from the dictionary onto the screen.

(15 marks)

---- END OF ASSIGNMENT ----