Jérémie Bourdon
Basic skills in programming
30 juin 2015

Programming in Python: Exercices

Basic exercises on strings

Exercise 1: Mimics the transcription phase by replacing all T by a U in a DNA sequence entered by the user.

Exercise 2: Write a function that returns the reverse complementary sequence of a DNA sequence provided as a parameter of the function.

Basic exercises on arrays

Exercise 3: A biologist have counted the number of individual of a given bird specie for 40 years. The data is provided by this table:

Année	Nb pas-sages	Année	Nb pas-sages	Année	Nb pas-sages	Année	Nb pas-sages
1	10	11	151	21	20	31	140
2	35	12	133	22	15	32	150
3	82	13	146	23	41	33	142
4	60	14	140	24	30	34	171
5	80	15	112	25	44	35	160
6	105	16	87	26	84	36	198
7	100	17	95	27	78	37	159
8	120	18	58	28	82	38	216
9	146	19	41	29	94	39	200
10	122	20	64	30	158	40	196

1) write a program that computes the year where the population is maximal
2) write a program that finds the pics of population.

Exercise 4: The correlation coefficient between two series is given by

$$r = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}$$

$$r = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}$$

Write a function that returns the correlation between two series provided as parameters of the function.

Ap ... ids in their faeces and (Y)
t

1	8	21	24	50	55	58	61	63	70	77	80	85	98	108	132	135	153
0	0	30	40	80	100	120	120	140	150	170	180						

x	0	0	30	40	80	100	120	120	140	150	170	180
y	0.04	0.02	0.00	0.02	0.12	0.08	0.06	0.15	0.16	0.11	0.17	0.12

Is the correlation between X and Y significant ?

$$Y'_i = \frac{Y_{i-1} + Y_i + Y_{i+1}}{3}$$

Simulations and plots

Exercise 5: The Fibonacci series (also known as « the rabbits » series) is defined is $F_0 = F_1 = 1$ and $F_n = F_{n-1} + F_{n-2}$, $n > 1$. Computes the 20 first terms of the Fibonacci series, store them in an array and

7	8	9	10	12	13	14	14	15	17	18	19	19	20	21	23
7	8	9	10	12	13	14	14	15	17	18	19	19	20	21	23
5.5	9.9	4.2	8.7	10.7	15.1	13.9	9.5	7.1	13.7	18.1	11.3	21.4	15.7	19.0	17.7

Exercise 6: The Lotka-Volterra model permits to reproduce some observed prey-predators phenomenon. More precisely, let X_n denotes the number of preys and Y_n denotes the number of predators. These series evolves as $X_{n+1} = V_x (1+X_n)$ and $Y_{n+1} = V_y (1+Y_n)$, with $V_x = a - b Y_n$ and $V_y = c X_n - d$.

Compute the series X and Y for a long period of time and plot the results. What can we conclude ?

Test it with a=0.01, b=0.02, c=0.03 and d=0.04….

Advanced exercises on strings

Exercise 5: The GC-content of a DNA string is given by the percentage of symbols in the string that are 'C' or 'G'. For example, the GC-content of "AGCTATAG" is 37.5%. Write a function that computes the GC content of a given DNA sequence.

Exercise 5: The distance between two DNA sequences can be measured in several way. The *Hamming distance* equals the number of positions for which the characters differs.

Write a function DistanceH.

The Levenshtein distance equals the minimal number of insertions, deletions and substitutions that are required to go from one sequence to another. Write a function DistanceL.

Example: DistanceH(ACTAATGA,ACAATGAC) = 5 and DistanceH(ACTAATGA,ACAATGAC) = 2….

Exercise 5: A k-mer is a string of length k. We define Count(Text, Pattern) as the number of times that a k-mer Pattern appears as a substring of Text. For example,

    Count(ACAACTATGCATACTATCGGGAACTATCCT,ACTAT)=3.

We note that Count(CGATATATCCATAG, ATA) is equal to 3 (not 2) since we should account for overlapping occurrences of Pattern in Text. We say that Pattern is a most frequent k-mer in Text if it maximizes Count(Text, Pattern) among all k-mers. For example, "ACTAT" is a most frequent 5-mer in "ACAACTATGCATCACTATCGGGAACTATCCT", and "ATA" is a most frequent 3-mer of "CGATATATCCATAG".

Frequent Words Problem

    Find the most frequent k-mers in a string.

    Given: A DNA string Text and an integer k.

    Return: All most frequent k-mers in Text (in any order).

Write the Count function and solve the Frequent Words Problem.

If it is not enough, go to http://rosalind.info and try to solve some problems…..