

Sheet 3

Free University of Berlin, WS 2007/08

Utz J. Pape · Martin Vingron

Assignments for the lecture PSSA
working period: 07.11.2007 - 14.11.2007

Exercise 6 (Dinucleotide Permutation Model). Rerun the analysis from exercise 1 while using a different sequence (and, therefore, a different di-nucleotide sequence model) for each run. Visualize, explain and interpret!!! (This means that I want to read what you did, why you did it, what you see (including pictures with explanations) and what you conclude from that.)

Exercise 7 (Exact Word Count Distribution). Repeat exercise 5 for the same background models and the same words and

1. compute the exact count distribution,
2. compare the results with the outcome using the independence model for the count distribution (old exercise). Here, comparison means that for each background model, both distributions are visualized together including a description of what you see).
3. Interpret! This means, again, what are your conclusions.
4. Analyze the running time.

Exercise 8 (Distances between two occurrences). Consider the words and background models from the last exercise:

1. visualize the 'distance between two occurrences' distribution for each word and each background model.
2. argue why the geometric distribution could also make sense.
3. select an appropriate parameter for the geometric distribution and compare the results with the first part of this exercise.
4. in which case (which might not be among the words you analyzed) do both distributions fit completely?