

Sheet 6

Free University of Berlin, WS 2007/08

Utz J. Pape · Martin Vingron

Assignments for the lecture PSSA
working period: 05.12.2007 - 12.12.2007

Exercise 14 (Number of Decompositions of a Sum). Given j clumps with total size of $l \geq j$. Prove that there are $\binom{l-1}{j-1}$ different configurations of clump sizes (e.g., one configuration is that $j-1$ clumps are of size 1 and one clump of size $l-j+1$).

Exercise 15 (Compound Poisson Approximation). Generalize and implement the compound Poisson distribution.

1. Generalize the formulae from the lecture to any word w . You can use the fifth chapter of Stephane Robin's book!
2. Add the Compound Poisson Approximation to the plots with the exact and the other approximate solutions for the four words and the three background models.
3. Also illustrate the error bounds - here, it might be better to use the cumulative distribution function.

Exercise 16 (p -value calculation). Compute the p -value for observing 35 hits for each of the words ACGT, ACAC, ATAT, and GGGG under the uniform background model using all known approximations (and error bounds if applicable) and the exact solution. Illustrate (including the error bounds) and interpret!