

Sheet 1

Free University of Berlin, WS 2007/08

Utz J. Pape · Martin Vingron

Assignments for the lecture PSSA
working period: 24.10.2007 - 31.10.2007

Exercise 1 (Dinucleotide Model). Download a 1kbp part of a real genome sequence (note down the organism, coordinates and download url). Count the dinucleotide occurrences for the first 100bp, 200bp, 500bp and the whole sequence. Generate for each of these sizes new sequences with the corresponding counts using the urn model. Track how many runs you need to succeed for each size. Interpret!

Exercise 2 (Variance). *changed!*

1. Proof $\text{Var } X = \mathbb{E} X^2 - \mathbb{E}^2 X$.
2. Compute $\text{Var } X$ where X is distributed like Bernoulli with $\mathbb{P}(X = 1) = p$.