

Übungsblatt 5

k-mer Identifikation

Deadline: 8.6.2022 um 20:00 MEZ

Bioinformatik für Biochemiestudierende

Dr. Florian Klimm

Sommersemester 2022

Aufgabe 5.1 Identische Zufallssequenzen (10%)

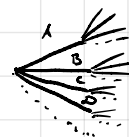
Wie lautet die Wahrscheinlichkeit, dass zwei Zufallssequenzen der Länge N mit einem Alphabet aus A Buchstaben identisch sind?

N_1 : A^N Möglichkeiten für Buchstabenfolge ; $\frac{1}{A^N}$ für 1 spezielle Sequenz

N_2 : A^N Möglichkeiten für Buchstabenfolge ; $\frac{1}{A^N}$ für 1 spezielle Sequenz

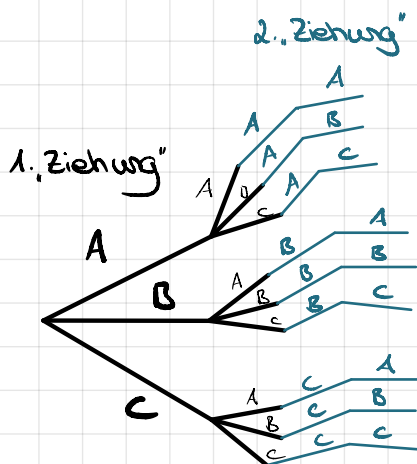
Kombination von N_1 & N_2 :

Nebenrechnung:



$$N=2 \rightarrow A^2$$

$$N=3 \rightarrow A^3$$



$$A = 3, \quad N = 2$$

\rightarrow beide Ziehungen hintereinander

A^N -Möglichkeiten

\rightarrow 1 Ziehung A^N

\rightarrow Wahrscheinlichkeit, dass beide Sequenzen gleich sind

$$\frac{A^N}{A^{2N}} = \frac{1}{A^N}$$

Aufgabe 5.2 Die Wahrscheinlichkeit von Zufallssequenzen (20 %)

Bestimmen Sie die Wahrscheinlichkeit $\Pr(N = 100, A = 2, \text{Pattern} = "01", t = 1)$.

$$\Pr(N, A, \text{Pattern}, t) = \frac{\binom{n+t}{t} \cdot A^n}{A^N} = \frac{\binom{N-t \cdot (k-1)}{t}}{A^{t \cdot k}}$$

$$n = N - t \cdot k = 100 - 1 \cdot 2 = 98$$

$$k = 2$$

$$\Pr(100, 2, "01", 1) = \frac{\binom{98+1}{1} 2^{98}}{2^{100}} = \underline{\underline{\frac{99}{4}}}$$