

⑤

Aufgabe 5.1 Identische Zufallssequenzen (10%)

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

Für jede Stelle ist die Wahrscheinlichkeit, dass diese identisch sind: $P_i = \frac{1}{A}$.

So ist für eine Sequenz mit N Stellen: $P_{i,gs} = \left(\frac{1}{A}\right)^N = \frac{1}{A^N}$

Aufgabe 5.2 Die Wahrscheinlichkeit von Zufallssequenzen (20%)

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

Aus der Vorlesung kennen wir eine Approximation dieser Wahrscheinlichkeit:

$$P_r(N, A, \text{Pattern}, t) = \frac{\binom{n+t}{t} \cdot A^t}{A^N}, \text{ wobei } n = N - t \text{ mit } t \text{ der Länge des Patterns}$$

Hier ist $k=2$, $t=1$, also $n = 100 - 2 = 98$. Damit

$$P_r(100, 2, "01", 1) = \binom{100}{1} 2^{-2} = 25. \text{ Für diesen Fall ist die Approximation nicht anwendbar.}$$

Dennach wird folgende Überlegung relevant: Damit in einer Sequenz mit $A=2$ ein 2-mer nur $t=1$ mal vorkommt,

müssen alle Einträge der Sequenz den gleichen Buchstaben repräsentieren b.z auf eine Stelle „0001000..“ oder „1110111..“

Die Wahrscheinlichkeit hierfür liegt bei $P_r(N=100, A=2, k=2, t=1) = P_i(N=99, A=2) \cdot \frac{1}{A} = \left(\frac{1}{2}\right)^{100} \approx 7,88 \cdot 10^{-31}$, also nahezu unmöglich.