**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?

An jeder Position tritt jeder Buchstabe mit der Wahrscheinlichkeit 1/A auf.

Egal, welcher Buchstabe die erste Sequenz hat, die Wahrscheinlichkeit, dass die zweite Sequenz den gleichen Buchstaben hat, ist auch 1/A.

Schauen wir uns jede Position in zwei Zufallssequenzen der Länge N unabhängig voneinander an, so ist die Wahrscheinlichkeit, dass alle Positionen übereinstimmen:

**Aufgabe 5.2 Die Wahrscheinlichkeit von Zufallssequenzen (20 %)**  
Bestimmen Sie die Wahrscheinlichkeit Pr(N = 100, A = 2, Pattern = ”01”, t = 1).

🡪 nicht über Wahrscheinlichkeit, sondern Gegenwahrscheinlichkeit:

In welchen Fällen tritt in der Sequenz kein „01“ 2-mer auf?

3 Fälle:

a) eine Sequenz, die nur aus Nullen besteht

b) eine Sequenz, die nur aus Einsen besteht

c) eine Sequenz, die aus 2 Teilen besteht: erst nur Einsen und dann nur Nullen; dabei kann der Wechsel von 0 nach 1 an jeder Stelle stattfinden  
🡪 a und b sind tatsächlich nur Spezialfälle von c, wenn Wechsel an der Position 0 und N zugelassen werden   
🡪 es ergeben sich N+1 mögliche Sequenzen der Art c

Die Anzahl möglicher Sequenzen, die N lang sind und aus einem Alphabet aus 2 Buchstaben bestehen sind N2. Daher ist die Wahrscheinlichkeit, kein 2-mer „01“ zu bekommen:

Es ergibt sich für die Wahrscheinlichkeit, mindestens eine „01“ zu bekommen: