

Exercise II, Sublinear Algorithms for Big Data Analysis 2024-2025

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students, and solve as many problems as you can. Problems marked (*) are more difficult, but also more rewarding. These problems have been taken from various sources on the Internet, too numerous to cite individually.

- 1** Show that for large enough integer n there exists a collection C of subsets of $[n] := \{1, 2, 3, \dots, n\}$ such that
 1. $|C| \geq 2^{cn}$ for some absolute constant $c > 0$;
 2. every element $S \in C$ is of size $n/4$;
 3. for every two $S_1 \neq S_2 \in C$, $|S_1 \cap S_2| \leq n/8$
- 2** Use the result of **1** to show that any **deterministic** algorithm that achieves a 1 ± 0.1 approximation to the number of distinct elements in a data stream must use $\Omega(n)$ space.