

Bucket Sort When You Know The Distribution.

David Ponnarovsky

January 20, 2023

Abstract

We propose a new simple construction based on Tanner Codes, which yields a good LDPC code with testability query complexity of $\Theta(n^{1-\varepsilon})$ for any $\varepsilon > 0$.

The problem. Let $f : [0, 1] \rightarrow [0, 1]$ a fixed distribution function. Write an algorithm that sort n draws $x_1 \dots x_n$ at linear expectation time.

Solution. We will define a partition of the input into a seira of n buckets $\mathcal{B} = \{B_k = [t_k, t_{k+1}] : k \in [n]\}$ such that $\Pr[x \in B_i] = \frac{1}{n}$ for any bucket.

Claim. The probability that the size of the i th bucket exceeds $t \in \mathbb{N}$ is bounded by: $\Pr[B_i \geq t] \leq \frac{e}{t^k}$ for every intrger $k \leq n$.

Proof. Let the X_{ij} be the indecator of the event that x_j belongs to B_i . Then we have:

$$\begin{aligned} \frac{1}{n} &= \Pr[x \in B_k] = f(t_{k+1}) - f(t_k) \\ \Rightarrow t_{k+1} &\leftarrow f^{-1}\left(\frac{1}{n} + f(t_k)\right) \end{aligned}$$