Bucket Sort When You Know The Distribution.

David Ponarovsky

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 $\Theta\left(n^{1-\varepsilon}\right)$ for any $\varepsilon > 0$.

The problem. Let $f : [0,1] \to [0,1]$ a fixed distribution function. Write an algorithm that sort n draws $x_1 ... x_n$ 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 $\mathbf{Pr}[x \in B_i] = \frac{1}{n}$ for any bucket.

$$\begin{aligned}\exp B_i^2 &= \exp\left(\sum_j X_{ij}\right)^2 = \exp \sum_{j,j'} X_{ij} X_{ij'} \\ &= \sum_{\{j,j'\} \exp X_{ij} \exp X_{ij'} = \sum_{\{j \neq j'\}} \exp X_{ij} \exp X_{ij'} + \sum_j \exp X_{ij} =} \\ &= \binom{n}{2} + 1 = O(1)\end{aligned}$$