Contribution ID: **110** Type: **not specified**

# Sample-based distance-approximation for subsequence-freeness

*Thursday, July 13, 2023 4:50 PM (20 minutes)*

Dana Ron and Omer Cohen Sidon

Abstract: In this work, we study the problem of approximating the distance to subsequence-freeness in the sample-based distribution-free model. For a given subsequence (word) $w = w_1 \dots w_k$, a sequence (text) $T = t_1 \dots t_n$ is said to contain $w$ if there exist indices $1 \leq i_1 < \cdots < i_k \leq n$ such that $t_{i_j} = w_j$ for every $1 \leq j \leq k$.
Otherwise, $T$ is $w$-free. Ron and Rosin (ACM TOCT 2022) showed that the number of samples both necessary and sufficient for one-sided error testing of subsequence-freeness in the sample-based distribution-free model is $\Theta(k/\epsilon)$.

Denoting by $Dist(T, w, p)$ the distance of $T$ to $w$-freeness under a distribution $p : [n] \to [0, 1]$, we are interested in obtaining an estimate $wDist$, such that $|wDist - Dist(T, w, p)| \leq \delta$ with probability at least $2/3$, for a given distance parameter $\delta$. Our main result is an algorithm whose sample complexity is $\tilde{O}(k^2/\delta^2)$. We first present an algorithm that works when the underlying distribution $p$ is uniform, and then show how it can be modified to work for any (unknown) distribution $p$. We also show that a quadratic dependence on $1/\delta$ is necessary.

**Presenter:** SIDON, Omer Cohen

**Session Classification:** Track A-3