Finding Monotone Patterns in Sublinear Time

Erik Waingarten (Columbia University)

Clément Canonne (Stanford University) Omri Ben-Eliezer (Tel-Aviv University) Shoham Letzter (ETH Zurich)



Testing Monotonicity of an Array: Sortedness



Given query access to an unknown $f: [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$:

- If f monotone, accept w.p. > 2/3;
- If f is ε -far from monotone, reject w.p. > 2/3;

Minimum query complexity in terms of n and ε ?

Testing Monotonicity of an Array: Sortedness



Given query access to an unknown $f: [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$:

- If f monotone, accept w.p. > 2/3;
- If f is ε -far from monotone, reject w.p. > 2/3;

Minimum query complexity in terms of n and ε ?

- If f is ε -far from monotone, find evidence:
 - $i, j \in [n]$ where i < j and f(i) > f(j).

Testing Monotonicity of an Array: Sortedness



Given query access to an unknown $f: [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$:

- If f monotone, accept w.p. > 2/3;
- If f is ε -far from monotone, reject w.p. > 2/3;

Minimum query complexity in terms of n and ε ?

- If f is ε -far from monotone, find evidence:
 - $i, j \in [n]$ where i < j and f(i) > f(j). one-sided error.

ε -Far-From-Monotone Sequences

• f is ε -far from monotone: for any monotone $g: [n] \to \mathbb{R}$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{f(i)\neq g(i)\}\geq \varepsilon.$$

ε -Far-From-Monotone Sequences

• f is ε -far from monotone: for any monotone $g \colon [n] \to \mathbb{R}$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{f(i)\neq g(i)\}\geq \varepsilon.$$

Lemma

Let $f: [n] \to \mathbb{R}$ be ε -far from monotone. There exists set of disjoint pairs,

$$T = \{(i, j) \in [n]^2 : i < j \text{ and } f(i) > f(j)\}$$

of size $|T| \ge \varepsilon n/2$.

Monotonicity Testing

Theorem (Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99)

There exists a non-adaptive, one-sided algorithm for testing monotonicity of $f: [n] \to \mathbb{R}$ making $O((\log n)/\varepsilon)$ queries.

Monotonicity Testing

Theorem (Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99)

There exists a non-adaptive, one-sided algorithm for testing monotonicity of $f : [n] \to \mathbb{R}$ making $O((\log n)/\varepsilon)$ queries.

• $\Omega((\log n)/\varepsilon)$ queries needed for non-adaptive algorithms.

Monotonicity Testing

Theorem (Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99)

There exists a non-adaptive, one-sided algorithm for testing monotonicity of $f: [n] \to \mathbb{R}$ making $O((\log n)/\varepsilon)$ queries.

- $\Omega((\log n)/\varepsilon)$ queries needed for non-adaptive algorithms.
- $\Omega((\log n)/\varepsilon)$ queries needed for adaptive algorithms too! [Fischer 04].

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

Definition

Let $k \in \mathbb{N}$ and $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of [k]. Given $f: [n] \to \mathbb{R}$, the k-tuple (i_1, \dots, i_k) has order pattern π if: $f(i_\ell) < f(i_m)$ whenever $\pi_\ell < \pi_m$.

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

Definition

Let $k \in \mathbb{N}$ and $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of [k]. Given $f: [n] \to \mathbb{R}$, the k-tuple (i_1, \dots, i_k) has order pattern π if: $f(i_\ell) < f(i_m)$ whenever $\pi_\ell < \pi_m$.

If $f: [n] \to \mathbb{R}$ is ε -far from π -free, then there exists $T \subset [n]^k$ of disjoint violating k-tuples (i_1, \ldots, i_k) with order pattern π of size at least $\varepsilon n/k$.

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

Definition

Let $k \in \mathbb{N}$ and $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of [k]. Given $f: [n] \to \mathbb{R}$, the k-tuple (i_1, \dots, i_k) has order pattern π if: $f(i_\ell) < f(i_m)$ whenever $\pi_\ell < \pi_m$.

If $f: [n] \to \mathbb{R}$ is ε -far from π -free, then there exists $T \subset [n]^k$ of disjoint violating k-tuples (i_1, \ldots, i_k) with order pattern π of size at least $\varepsilon n/k$.

For fixed k and π , query complexity of testing π -freeness?

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

Definition

Let $k \in \mathbb{N}$ and $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of [k]. Given $f: [n] \to \mathbb{R}$, the k-tuple (i_1, \dots, i_k) has order pattern π if: $f(i_\ell) < f(i_m)$ whenever $\pi_\ell < \pi_m$.

If $f: [n] \to \mathbb{R}$ is ε -far from π -free, then there exists $T \subset [n]^k$ of disjoint violating k-tuples (i_1, \ldots, i_k) with order pattern π of size at least $\varepsilon n/k$.

For fixed k and π , query complexity of testing π -freeness?

• Some sublinear in *n* upper bounds for general π .

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

Definition

Let $k \in \mathbb{N}$ and $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of [k]. Given $f: [n] \to \mathbb{R}$, the k-tuple (i_1, \dots, i_k) has order pattern π if: $f(i_\ell) < f(i_m)$ whenever $\pi_\ell < \pi_m$.

If $f: [n] \to \mathbb{R}$ is ε -far from π -free, then there exists $T \subset [n]^k$ of disjoint violating k-tuples (i_1, \ldots, i_k) with order pattern π of size at least $\varepsilon n/k$.

For fixed k and π , query complexity of testing π -freeness?

- Some sublinear in *n* upper bounds for general π .
- $\pi = (132)$ requires $\Omega(\sqrt{n})$ queries for non-adaptive, one-sided algorithms.

[Newman, Rabinovich, Rajendraprasad, Sohler 17]

Definition

Let $k \in \mathbb{N}$ and $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of [k]. Given $f: [n] \to \mathbb{R}$, the k-tuple (i_1, \dots, i_k) has order pattern π if: $f(i_\ell) < f(i_m)$ whenever $\pi_\ell < \pi_m$.

If $f: [n] \to \mathbb{R}$ is ε -far from π -free, then there exists $T \subset [n]^k$ of disjoint violating k-tuples (i_1, \ldots, i_k) with order pattern π of size at least $\varepsilon n/k$.

For fixed k and π , query complexity of testing π -freeness?

- Some sublinear in *n* upper bounds for general π .
- $\pi = (132)$ requires $\Omega(\sqrt{n})$ queries for non-adaptive, one-sided algorithms.
- [Ben-Eliezer, Canonne 18] Many π have complexity a $n^{1-1/(k-\Theta(1))}$.

Given query access to $f : [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$ where:

Given query access to $f \colon [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$ where:

• There exists $T \subset [n]^k$ of disjoint violating k-tuples

$$\mathcal{T} = \{(i_1,\ldots,i_k): i_1 < \cdots < i_k ext{ and } f(i_1) < \cdots < f(i_k)\}$$

of size $|T| \ge \varepsilon n/k$.

Given query access to $f \colon [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$ where:

• There exists $T \subset [n]^k$ of disjoint violating k-tuples

$$\mathcal{T} = \{(i_1,\ldots,i_k): i_1 < \cdots < i_k ext{ and } f(i_1) < \cdots < f(i_k)\}$$

of size $|T| \ge \varepsilon n/k$.

• Find $i_1 < \cdots < i_k$ where $f(i_1) < \cdots < f(i_k)$.

Given query access to $f \colon [n] \to \mathbb{R}$ and a parameter $\varepsilon > 0$ where:

• There exists $T \subset [n]^k$ of disjoint violating k-tuples

$$T = \{(i_1, \ldots, i_k) : i_1 < \cdots < i_k \text{ and } f(i_1) < \cdots < f(i_k)\}$$

of size
$$|T| \ge \varepsilon n/k$$
.

• Find
$$i_1 < \cdots < i_k$$
 where $f(i_1) < \cdots < f(i_k)$.

Theorem (NRRS17)

There is a non-adaptive algorithm with query complexity $((\log n)/\varepsilon)^{O(k^2)}$.

Theorem (Upper bound)

For $k \in \mathbb{N}$, there exists a non-adaptive algorithm with query complexity $(\log n)^{\lfloor \log_2 k \rfloor} \cdot \operatorname{poly}(1/\varepsilon).$

Theorem (Upper bound)

For $k \in \mathbb{N}$, there exists a non-adaptive algorithm with query complexity $(\log n)^{\lfloor \log_2 k \rfloor} \cdot \operatorname{poly}(1/\varepsilon).$

Theorem (Lower bound)

Any non-adaptive algorithm needs to make $\Omega\left((\log n)^{\lfloor \log_2 k \rfloor}\right)$ queries.

Let $f : [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with $T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

Sample ℓ ~ [n] uniformly and repeat the following for t iterations:
Sample s ~ {0,..., log₂ n} and i ~ [ℓ - 2^s, ℓ] and query f(i).

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

- ▶ Sample $s \sim \{0, ..., \log_2 n\}$ and $i \sim [\ell 2^s, \ell]$ and query f(i).
- ▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

- ▶ Sample $s \sim \{0, \dots, \log_2 n\}$ and $i \sim [\ell 2^s, \ell]$ and query f(i).
- ▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

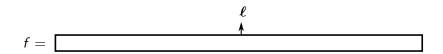
$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

▶ Sample
$$s \sim \{0, \dots, \log_2 n\}$$
 and $i \sim [\ell - 2^s, \ell]$ and query $f(i)$.

▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.



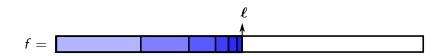
Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

- ▶ Sample $s \sim \{0, \dots, \log_2 n\}$ and $i \sim [\ell 2^s, \ell]$ and query f(i).
- ▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.



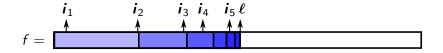
Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

- ▶ Sample $s \sim \{0, \dots, \log_2 n\}$ and $i \sim [\ell 2^s, \ell]$ and query f(i).
- ▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.



Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

▶ Sample $s \sim \{0, \dots, \log_2 n\}$ and $i \sim [\ell - 2^s, \ell]$ and query f(i).

▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.

Let $f: [n] \to \mathbb{R}$ and disjoint subset of pairs T of size $\varepsilon n/2$ with

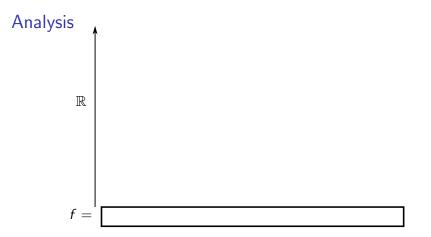
$$T = \{(i,j) \in [n]^2 : i < j \text{ and } f(i) < f(j)\}.$$

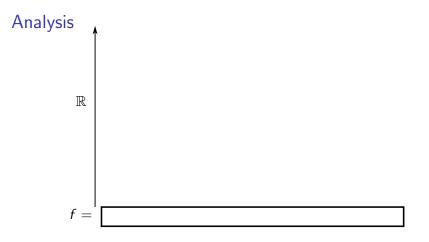
[Ergün, Kannan, Kumar, Rubinfeld, Viswanathan 99] Find increasing pair:

• Sample $\ell \sim [n]$ uniformly and repeat the following for t iterations:

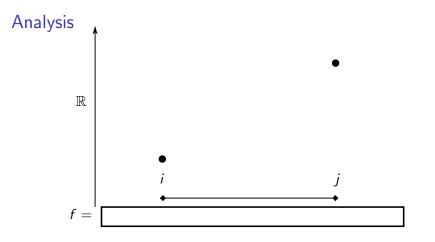
- ▶ Sample $s \sim \{0, \dots, \log_2 n\}$ and $i \sim [\ell 2^s, \ell]$ and query f(i).
- ▶ Sample $\boldsymbol{s} \sim \{0, \dots, \log_2 n\}$ and $\boldsymbol{j} \sim [\boldsymbol{\ell}, \boldsymbol{\ell} + 2^{\boldsymbol{s}}]$ and query $f(\boldsymbol{j})$.

$$f = \begin{bmatrix} \mathbf{i}_1 & \mathbf{i}_2 & \mathbf{i}_3 & \mathbf{i}_4 & \mathbf{i}_5 \ell & \mathbf{j}_1 & \mathbf{j}_2 & \mathbf{j}_3 & \mathbf{j}_4 & \mathbf{j}_5 \\ \uparrow & \uparrow \\ \hline \end{bmatrix}$$

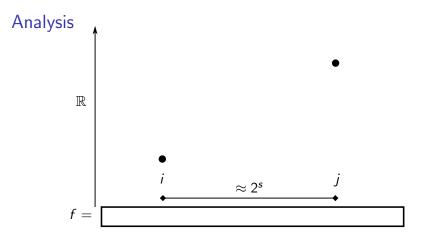




• Let T be set of monotone pairs.

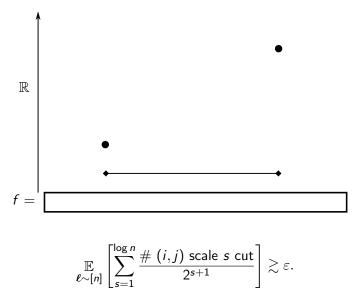


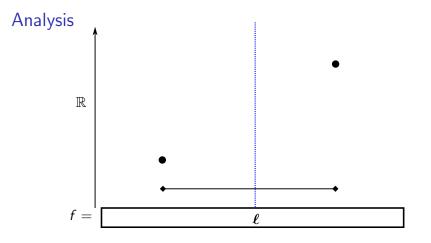
- Let T be set of monotone pairs.
- $(i,j) \in T$.



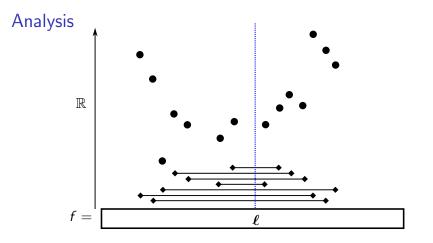
- Let T be set of monotone pairs.
- $(i,j) \in T$. (i,j) scale s when $j i \approx 2^s$. $\Pr_{\ell \sim [n]} [i \le \ell \le j] \approx \frac{2^s}{n}.$

Analysis

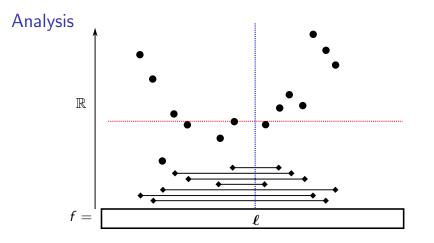




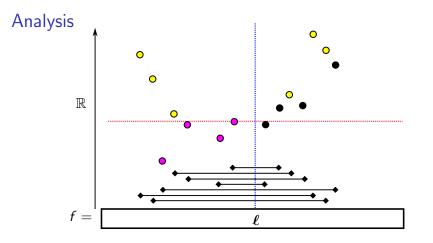
$$\sum_{s=1}^{\log n} \frac{\# \ (i,j) \text{ scale } s \text{ cut}}{2^{s+1}} = \sum_{s=1}^{\log n} (\text{density of } T \text{ at } 2^{s+1}) \gtrsim \varepsilon.$$



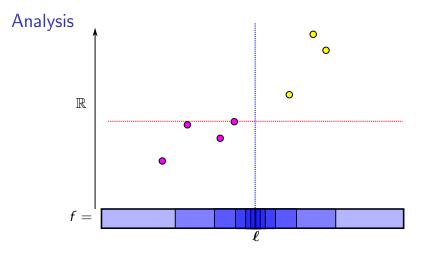
$$\sum_{s=1}^{\log n} \frac{\# \ (i,j) \text{ scale } s \text{ cut}}{2^{s+1}} = \sum_{s=1}^{\log n} (\text{density of } T \text{ at } 2^{s+1}) \gtrsim \varepsilon.$$



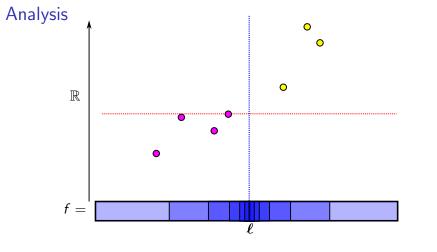
$$\sum_{s=1}^{\log n} \frac{\# \ (i,j) \text{ scale } s \text{ cut}}{2^{s+1}} = \sum_{s=1}^{\log n} (\text{density of } T \text{ at } 2^{s+1}) \gtrsim \varepsilon.$$



$$\sum_{s=1}^{\log n} \frac{\# \ (i,j) \text{ scale } s \text{ cut}}{2^{s+1}} = \sum_{s=1}^{\log n} (\text{density of } T \text{ at } 2^{s+1}) \gtrsim \varepsilon.$$



$$\sum_{s=1}^{\log n} \frac{\# \ (i,j) \text{ scale } s \text{ cut}}{2^{s+1}} = \sum_{s=1}^{\log n} (\text{density of } T \text{ at } 2^{s+1}) \gtrsim \varepsilon.$$



• Density δ_s^- and δ_s^+ : $\sum_{s=1}^{\log n} \delta_s^- \ge \varepsilon$ and $\sum_{s=1}^{\log n} \delta_s^+ \ge \varepsilon$. Pr i_{1,\dots,i_t} [avoid pair] $\le \left(\frac{1}{\log n} \sum_{s=1}^{\log n} (1 - \delta_s^-)\right)^t + \left(\frac{1}{\log n} \sum_{s=1}^{\log n} (1 - \delta_s^+)\right)^t \le \frac{1}{3}$.

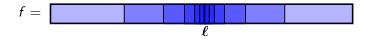
New algorithm for finding $(12 \dots k)$ patterns:

• Round 1: Sample $O(1/\varepsilon)$ indices from [n], include them in a set **A**.

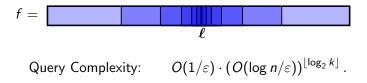
- Round 1: Sample $O(1/\varepsilon)$ indices from [n], include them in a set **A**.
- Round r, 2 ≤ r ≤ ⌊log₂ k⌋ + 1: For each i ∈ A and s ∈ {1,..., log n}, sample O(1/ε) indices from [i − 2^s, i + 2^s] and include them in A.

- Round 1: Sample $O(1/\varepsilon)$ indices from [n], include them in a set **A**.
- Round r, 2 ≤ r ≤ ⌊log₂ k⌋ + 1: For each i ∈ A and s ∈ {1,..., log n}, sample O(1/ε) indices from [i − 2^s, i + 2^s] and include them in A.
- Query f(i) for all $i \in \mathbf{A}$.

- Round 1: Sample $O(1/\varepsilon)$ indices from [n], include them in a set **A**.
- Round r, 2 ≤ r ≤ ⌊log₂ k⌋ + 1: For each i ∈ A and s ∈ {1,..., log n}, sample O(1/ε) indices from [i − 2^s, i + 2^s] and include them in A.
- Query f(i) for all $i \in \mathbf{A}$.

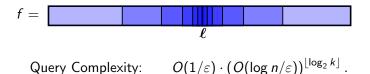


- Round 1: Sample $O(1/\varepsilon)$ indices from [n], include them in a set **A**.
- Round r, 2 ≤ r ≤ ⌊log₂ k⌋ + 1: For each i ∈ A and s ∈ {1,..., log n}, sample O(1/ε) indices from [i − 2^s, i + 2^s] and include them in A.
- Query f(i) for all $i \in \mathbf{A}$.



New algorithm for finding $(12 \dots k)$ patterns:

- Round 1: Sample $O(1/\varepsilon)$ indices from [n], include them in a set **A**.
- Round r, 2 ≤ r ≤ ⌊log₂ k⌋ + 1: For each i ∈ A and s ∈ {1,..., log n}, sample O(1/ε) indices from [i − 2^s, i + 2^s] and include them in A.
- Query f(i) for all $i \in \mathbf{A}$.



Theorem

Suppose $f : [n] \to \mathbb{R}$ contains $\varepsilon n/k$ disjoint $(12 \dots k)$ -patterns, algorithm finds one $w.p \ge 2/3$.

• More efficient tests for *two-sided* error testing for $(12 \dots k)$ -freeness.

- More efficient tests for *two-sided* error testing for $(12 \dots k)$ -freeness.
- Adaptive algorithms for finding π -patterns?

- More efficient tests for *two-sided* error testing for $(12 \dots k)$ -freeness.
- Adaptive algorithms for finding π -patterns?
 - ▶ Possibly poly(log n) adaptive queries for any pattern π?

- More efficient tests for *two-sided* error testing for $(12 \dots k)$ -freeness.
- Adaptive algorithms for finding π -patterns?
 - ▶ Possibly poly(log n) adaptive queries for any pattern π?

Thanks!