Title: Quantum algorithm for Statistical Difference problem

Date: Feb 18, 2009 04:00 PM

URL: http://pirsa.org/09020008

Abstract: Suppose we are given two probability distributions on some N-element set. How many samples do we need to test whether the two distributions are close or far from each other in the L_1 norm? This problem known as Statistical Difference has been extensively studied during the last years in the field of property testing. I will describe quantum algorithms for Statistical Difference problem that provide a polynomial speed up in terms of the query complexity compared to the known classical lower bounds. Specifically, I will assume that each distribution can be generated by querying an oracle function on a random uniformly distributed input string. It will be shown that testing whether distributions are orthogonal requires approximately N^{1/2} queries classically and approximately N^{1/3} queries quantumly. Testing whether distributions are close requires approximately N^{2/3} queries classically and O(N^{1/2}) queries quantumly. This is a joint work with Aram Harrow (University of Bristol) and Avinatan Hassidim (The Hebrew University).

Quantum algorithms for testing properties of probability distributions

Sergey Bravyi *(IBM Watson)*
Aram Harrow *(University of Bristol)*
Avinatan Hassidim *(The Hebrew University)*

Perimeter Institute
02/18/09

p, q — unknown probability distributions on $\{1, 2, \ldots, N\}$

Oracles O_p, O_q return samples from p, q

ϵ — constant precision parameter

p, q — unknown probability distributions on $\{1, 2, \ldots, N\}$

Oracles O_p, O_q return samples from p, q

ϵ — constant precision parameter

Property	Accept	Reject
Uniformity	$p = \frac{I}{N}$	$\lVert p - \frac{I}{N} \rVert_1 \geq \epsilon$
Closeness	$p = q$	$\lVert p - q \rVert_1 \geq \epsilon$
Orthogonality	p and q have disjoint support, $\lVert p - q \rVert_1 = 2$	p and q have constant overlap, $\lVert p - q \rVert_1 \leq 2 - \epsilon$

p, q — unknown probability distributions on $\{1, 2, \ldots, N\}$

Oracles O_p, O_q return samples from p, q

ϵ — constant precision parameter

Property	Accept	Reject
Uniformity	$p = \frac{I}{N}$	$\|p - \frac{I}{N}\|_1 \geq \epsilon$
Closeness	$p = q$	$\|p - q\|_1 \geq \epsilon$
Orthogonality	p and q have disjoint support, $\|p - q\|_1 = 2$	p and q have constant overlap, $\|p - q\|_1 \leq 2 - \epsilon$

How many samples do we need to test a property?

Motivation

Testing a property often requires only sublinear number of samples, e.g. $N^{2/3}$ or $N^{1/2}$.

Motivation

Testing a property often requires only sublinear number of samples, e.g. $N^{2/3}$ or $N^{1/2}$.

Uniformity: testing whether a random walk on a black-box graph is rapidly mixing

Motivation

Testing a property often requires only sublinear number of samples, e.g. $N^{2/3}$ or $N^{1/2}$.

Uniformity: testing whether a random walk on a black-box graph is rapidly mixing

Closeness: testing whether statistical experimental data agree with theoretical predictions. Testing whether a Markov chain is rapidly mixing.

Motivation

Testing a property often requires only sublinear number of samples, e.g. $N^{2/3}$ or $N^{1/2}$.

Uniformity: testing whether a random walk on a black-box graph is rapidly mixing

Closeness: testing whether statistical experimental data agree with theoretical predictions. Testing whether a Markov chain is rapidly mixing.

Orthogonality: SZK-complete problem if the oracles have explicit description [Vadhan 97].

Outline

(1) Statement of the problem and main results

(2) Classical lower bounds (P. Valiant 2008)

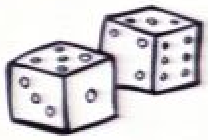
(3) Testing orthogonality and Collision Finding problem

(4) Testing closeness

(5) Testing uniformity

(6) Conclusions

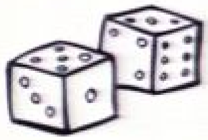
Statement of the problem



n-bit string → **Classical deterministic oracle O_p** → sample $i \in [N]$

$$p_i = \frac{\#\text{ inputs leading to an output } i}{\#\text{ inputs}}$$

Statement of the problem



n-bit string → Classical deterministic oracle O_p → sample $i \in [N]$

$$p_i = \frac{\#\ \text{inputs leading to an output}\ i}{\#\ \text{inputs}}$$

Quantum oracle: $\hat{O}_p : |x\rangle \otimes |0\rangle = |x\rangle \otimes |O_p(x)\rangle$

Statement of the problem

m-bit string → Classical deterministic oracle O_p → sample $i \in [N]$

$$p_i = \frac{\#\ \text{inputs leading to an output } i}{\#\ \text{inputs}}$$

Quantum oracle: $\hat{O}_p : |x\rangle \otimes |0\rangle = |x\rangle \otimes |O_p(x)\rangle$

Property tester:

Input: m, N, ϵ, access to a (quantum) oracle

Output: Accept or Reject

constant error probability, constant precision ϵ

Previous work and main results

Property	Cl. Upper	Cl. Lower	Q. Upper	Q. Lower
Uniformity	$\tilde{O}(N^{2/3})$	$\Omega(N^{1/2})$	$O(N^{1/3})$	?
Closeness	$\tilde{O}(N^{2/3})$	$\Omega(N^{2/3})$	$O(N^{1/2})$	?
Orthogonality	$O(N^{1/2})$	$\Omega(N^{1/2})$	$O(N^{1/3})$	$\Omega(N^{1/3})$

Relevant papers:

1] Batu, Fortnov et al, Testing that distributions are close, FOCS 2000

2] Valiant, Testing symmetric properties of distributions, STOC 2008

3] Goldreich and Ron, A sublinear bipartiteness tester ..., STOC 1998

Classical lower bounds (Valiant 2008)

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision of order k: Some element i appears in X exactly k times

Classical lower bounds (Valiant 2008)

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision of order k: Some element i appears in X exactly k times

Example: $X = (1, 3, 1, 2, 3, 1, 2, 4)$

1 collision of order 1 ($i = 4$)

2 collisions of order 2 ($i = 2$ and $i = 3$)

1 collision of order 3 ($i = 1$)

Classical lower bounds (Valiant 2008)

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision of order k: Some element i appears in X exactly k times

Example: $X = (1, 3, 1, 2, 3, 1, 2, 4)$

1 collision of order 1 ($i = 4$)

2 collisions of order 2 ($i = 2$ and $i = 3$)

1 collision of order 3 ($i = 1$)

c_k=# collisions of order k

Fingerprint of X: $c = (c_1, c_2, \ldots, c_M)$

Classical lower bounds (Valiant 2008)

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision of order k: Some element i appears in X exactly k times

Example: $X = (1, 3, 1, 2, 3, 1, 2, 4)$

1 collision of order 1 ($i = 4$)

2 collisions of order 2 ($i = 2$ and $i = 3$)

1 collision of order 3 ($i = 1$)

c_k=# collisions of order k

Fingerprint of X: $c = (c_1, c_2, \ldots, c_M)$

Example above: $c = (1, 2, 1, 0, 0, 0, 0, 0)$

Classical lower bounds (Valiant 2008)

(1) A fingerprint contains all relevant information for testing symmetric properties (invariant under relabeling of elements)

Classical lower bounds (Valiant 2008)

(1) A fingerprint contains all relevant information for testing symmetric properties (invariant under relabeling of elements)

Corollary: let D_p^M be a probability distribution of fingerprints. If a tester is supposed to accept p and reject q but

$$\|D_p^M - D_q^M\|_1 \ll \epsilon$$

then M samples is not enough to test a property.

Classical lower bounds (Valiant 2008)

2) Wishful Thinking Theorem (simplified version)

Suppose $\|p\|_\infty$ and $\|q\|_\infty$ are small compared with $1/M$. Then

$$\|D_p^M - D_q^M\|_1 \leq O(1) \sum_{k=2}^{\infty} M^{k/2} \frac{|\theta_k(p) - \theta_k(q)|}{\sqrt{\max\{\theta_k(p), \theta_k(q)\}}}$$

where $\theta_k(p) = \sum_{i=1}^{N} p_i^k$ is the k-th moment of p

3) Simple generalization to properties that involve two distributions, such as orthogonality and closeness.

Testing orthogonality using $O(N^{1/3})$ queries

Accept if p, q have disjoint support, $\|p - q\|_1 = 2$

Reject if p, q have constant overlap, $\|p - q\|_1 \leq 2 - \epsilon$

Wishful Thinking Theorem provides classical lower bounds

Uniformity testing: $\Omega(N^{1/2})$

Orthogonality testing: $\Omega(N^{1/2})$

Closeness testing: $\Omega(N^{2/3})$

More general problem: estimating $\|p-q\|_1$ with a constant precision. It requires $\Omega(N^{1-o(1)})$ queries.

Testing orthogonality using $O(N^{1/3})$ queries

Accept if p, q have disjoint support, $\|p - q\|_1 = 2$

Reject if p, q have constant overlap, $\|p - q\|_1 \leq 2 - \epsilon$

Testing orthogonality using $O(N^{1/3})$ queries

Accept if p, q have disjoint support, $\|p - q\|_1 = 2$
Reject if p, q have constant overlap, $\|p - q\|_1 \leq 2 - \epsilon$

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision probability: $$r = \sum_{i \in X} q_i$$

Testing orthogonality using $O(N^{1/3})$ queries

Accept if p, q have disjoint support, $\|p - q\|_1 = 2$
Reject if p, q have constant overlap, $\|p - q\|_1 \leq 2 - \epsilon$

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision probability: $$r = \sum_{i \in X} q_i$$

Basic intuition:

$p \perp q$ implies $r = 0$ with probability 1 (no collisions)

Testing orthogonality using $O(N^{1/3})$ queries

Accept if p, q have disjoint support, $\|p - q\|_1 = 2$

Reject if p, q have constant overlap, $\|p - q\|_1 \leq 2 - \epsilon$

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision probability: $\quad r = \sum_{i \in X} q_i$

Basic intuition:

$p \perp q$ implies $r = 0$ with probability 1 (no collisions)

$\|p - q\|_1 \leq 2 - \epsilon$ implies $r \geq const \cdot \frac{M}{N}$ w.h.p.

Testing orthogonality using $O(N^{1/3})$ queries

Accept if p, q have disjoint support, $\|p-q\|_1 = 2$
Reject if p, q have constant overlap, $\|p-q\|_1 \le 2-\epsilon$

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision probability: $$r = \sum_{i \in X} q_i$$

Basic intuition:

$p \perp q$ implies $r = 0$ with probability 1 (no collisions)

$\|p-q\|_1 \le 2-\epsilon$ implies $r \ge const \cdot \frac{M}{N}$ w.h.p.

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision probability: $r = \sum_{i \in X} q_i$

Large deviation bound:

Suppose $\|p - q\|_1 \le 2 - \epsilon$ and $32\epsilon^{-1} \le M \le N/2$. Then

$$\Pr\left[r \ge \frac{\epsilon^2}{256} \frac{M}{N} \right] \ge \frac{1}{3}.$$

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

Collision probability: $r = \sum_{i \in X} q_i$

Large deviation bound:

Suppose $\|p - q\|_1 \le 2 - \epsilon$ and $32\epsilon^{-1} \le M \le N/2$. Then

$$\Pr\left[r \ge \frac{\epsilon^2}{256} \frac{M}{N} \right] \ge \frac{1}{3}.$$

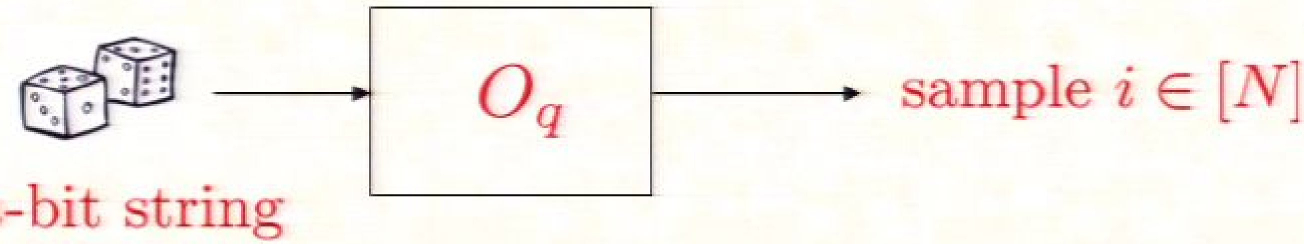
Remark: in the regime $M \sim N^{1/3}$ the standard deviation of r is much larger than the expectation value. One cannot use Chebyshev inequality.

Finding a collision using [BHT 97]

1. Let $X = (i_1, \ldots, i_M)$ be a list of M samples drawn from p

Finding a collision using [BHT 97]

1. Let $X = (i_1, \ldots, i_M)$ be a list of M samples drawn from p

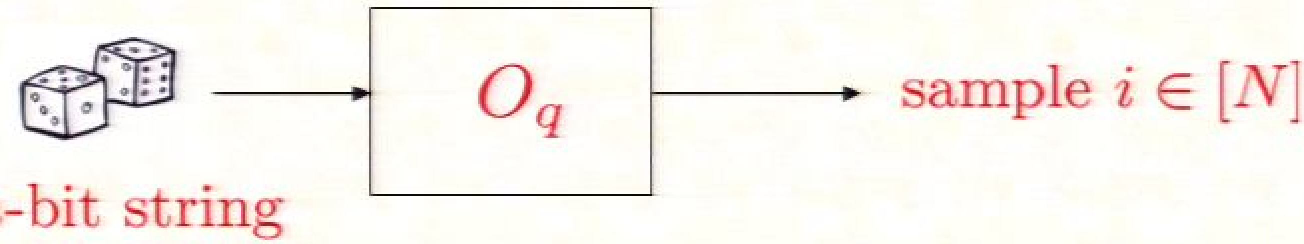


$$q_i = \frac{\#\text{ inputs leading to an output } i}{\#\text{ inputs}}$$

2. Mark all input strings y such that $O_q(y) \in X$
   Collision probability r = fraction of marked strings

Finding a collision using [BHT 97]

1. Let $X = (i_1, \ldots, i_M)$ be a list of M samples drawn from p



$$q_i = \frac{\#\text{ inputs leading to an output } i}{\#\text{ inputs}}$$

2. Mark all input strings y such that $O_q(y) \in X$
   Collision probability $r =$ fraction of marked strings

3. Assuming a lower bound $r \geq r_{min} \sim M/N$ find a marked string using the Grover search.

Finding a collision using [BHT 97]

1. Let $X = (i_1, \ldots, i_M)$ be a list of M samples drawn from p

$$q_i = \frac{\#\text{ inputs leading to an output } i}{\#\text{ inputs}}$$

2. Mark all input strings y such that $O_q(y) \in X$
   Collision probability $r =$ fraction of marked strings
3. Assuming a lower bound $r \geq r_{min} \sim M/N$ find a marked string using the Grover search.
4. If a marked string is found, reject. Otherwise accept.

Finding a collision using [BHT 97]

1. Let $X = (i_1, \ldots, i_M)$ be a list of M samples drawn from p

$$q_i = \frac{\#\text{ inputs leading to an output } i}{\#\text{ inputs}}$$

2. Mark all input strings y such that $O_q(y) \in X$
   Collision probability r = fraction of marked strings

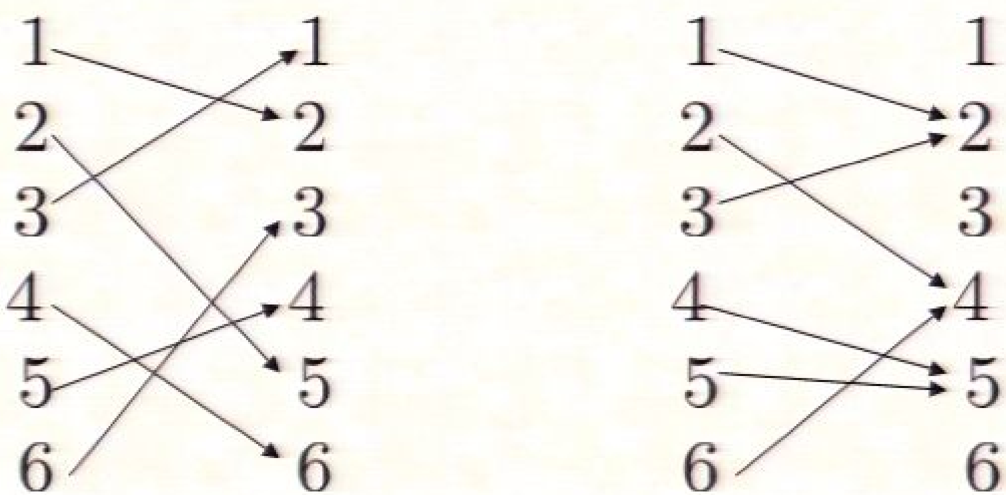
3. Assuming a lower bound $r \geq r_{min} \sim M/N$ find a marked string using the Grover search.

4. If a marked string is found, reject. Otherwise accept.

$$\#\text{ queries} = M + O\left(\sqrt{\frac{1}{r_{min}}}\right) = M + O\left(\sqrt{\frac{N}{M}}\right) = O(N^{1/3}).$$

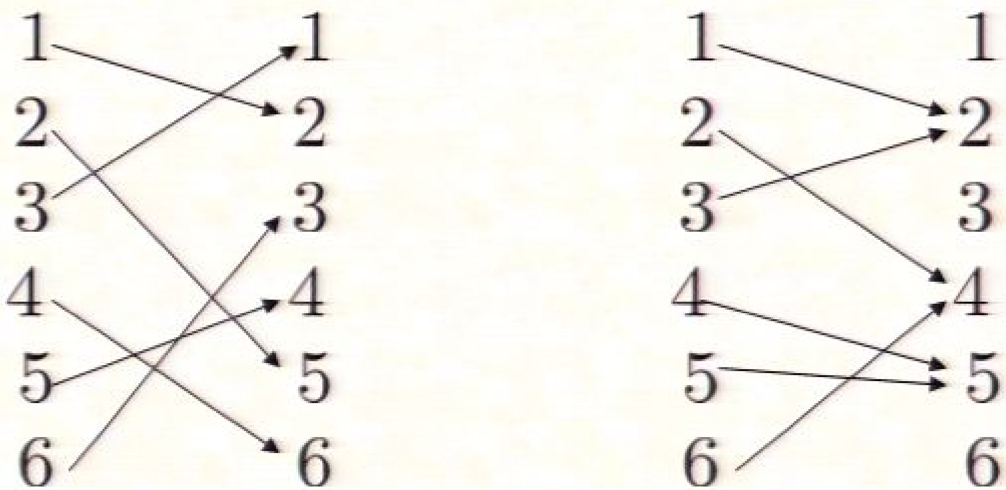
Collision Finding Problem

Decide whether an oracle function $F : [N] \rightarrow [N]$ is one-to-one or two-to-one.



Collision Finding Problem

Decide whether an oracle function $F : [N] \to [N]$ is one-to-one or two-to-one.

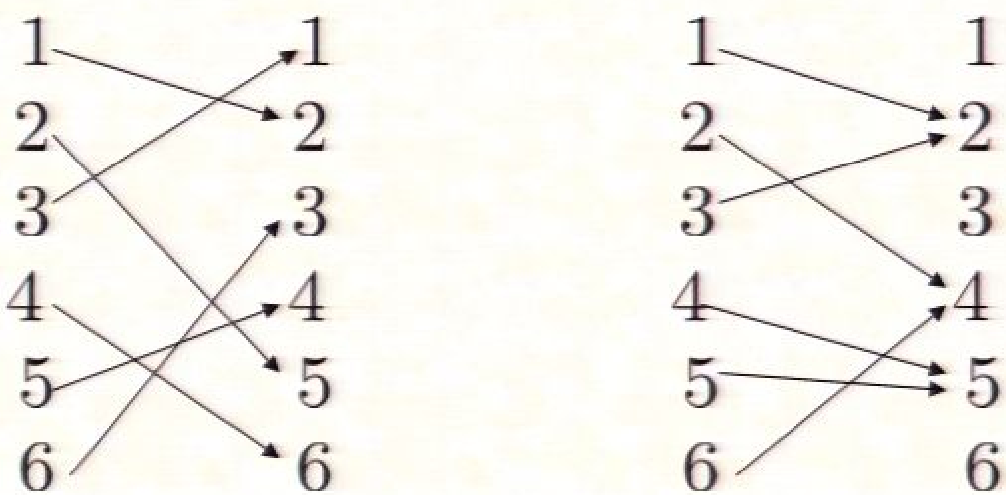


Brassard, Hoyer, Tapp 98: $O(N^{1/3})$ algorithm

Aaronson and Shi 04: $\Omega(N^{1/3})$ lower bound

Collision Finding Problem

Decide whether an oracle function $F : [N] \rightarrow [N]$ is one-to-one or two-to-one.



Brassard, Hoyer, Tapp 98: $O(N^{1/3})$ algorithm

Aaronson and Shi 04: $\Omega(N^{1/3})$ lower bound

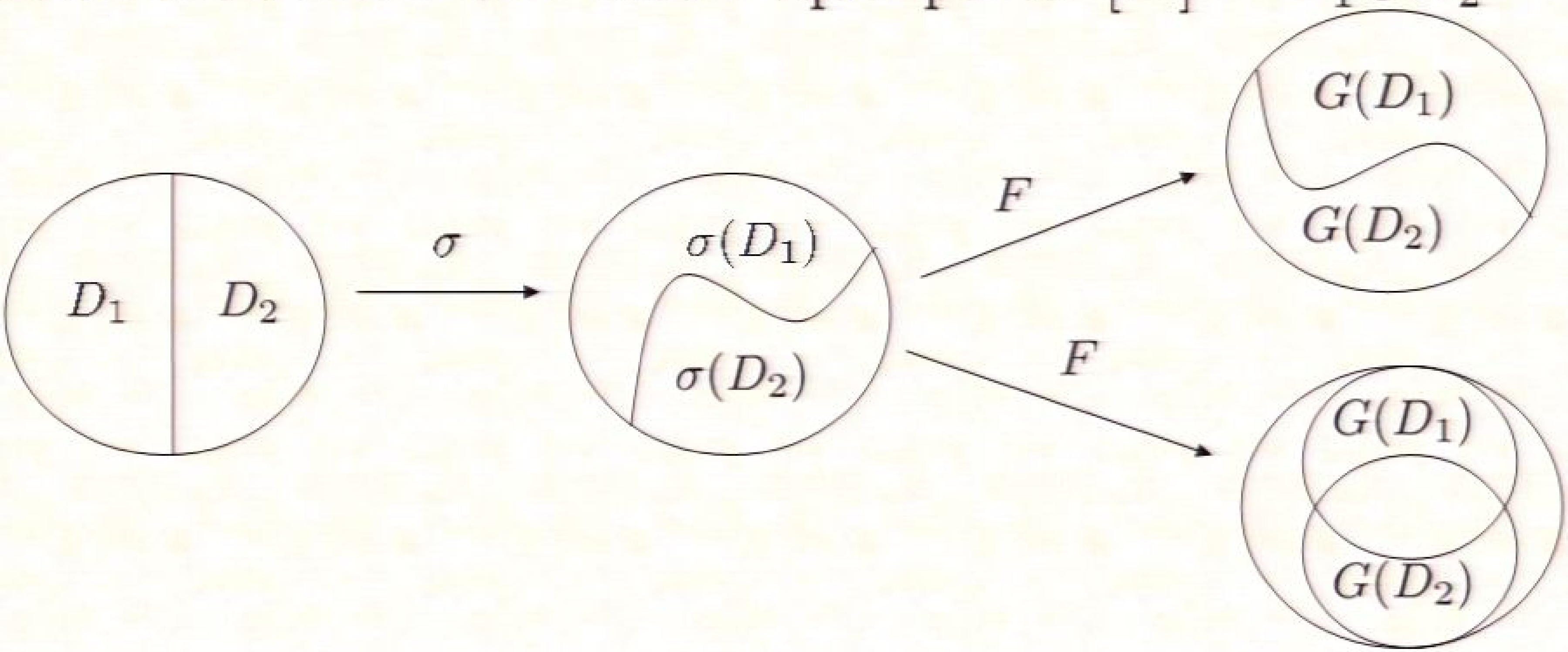
Simple observation: Collision Finding Problem is a special case of orthogonality testing.

Lower bound $\Omega(N^{1/3})$ for orthogonality testing

Choose a random permutation $\sigma \in S_N$

Define a new oracle $G = F \circ \sigma$

Partition the domain of F into 2 equal parts: $[N] = D_1 \cup D_2$

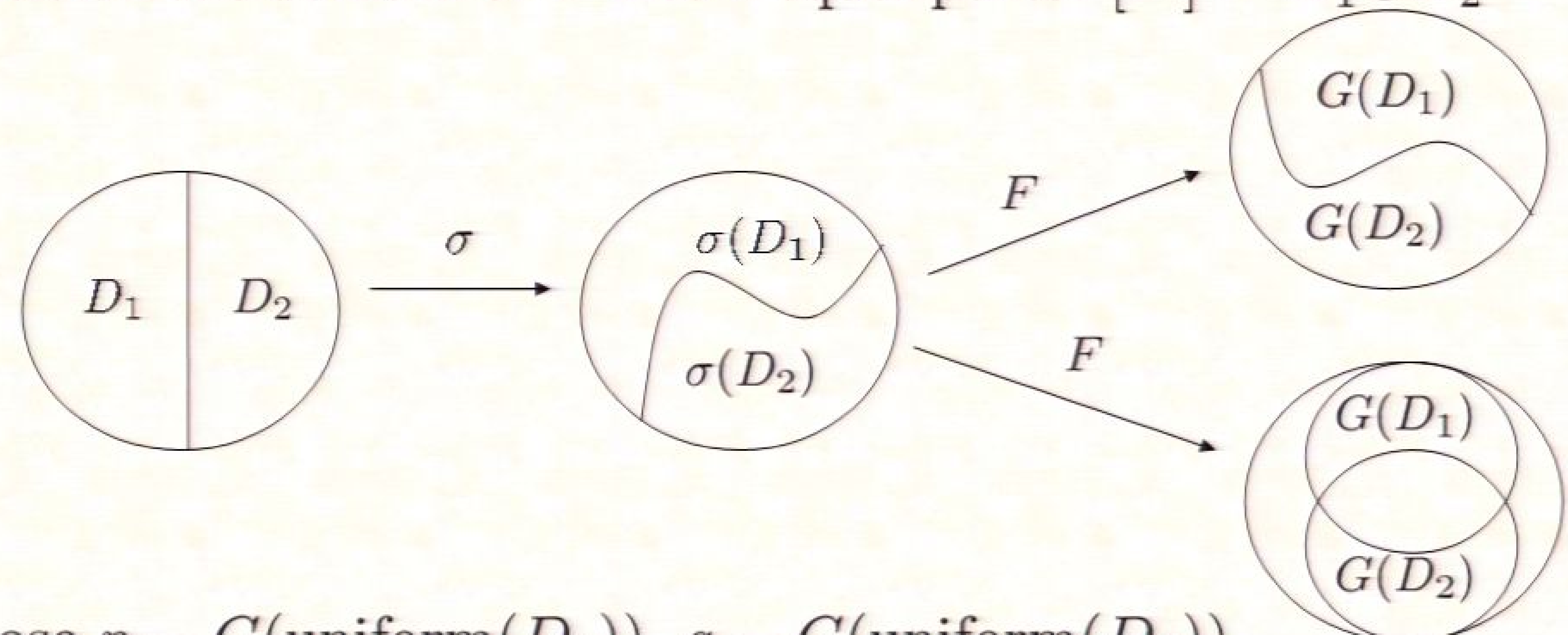


Lower bound $\Omega(N^{1/3})$ for orthogonality testing

Choose a random permutation $\sigma \in S_N$

Define a new oracle $G = F \circ \sigma$

Partition the domain of F into 2 equal parts: $[N] = D_1 \cup D_2$



Choose $p = G(\text{uniform}(D_1))$, $q = G(\text{uniform}(D_2))$

If F is one-to-one then $p \perp q$.

If F is two-to-one then $\Pr[\|p - q\|_1 \leq 7/8] \geq 1/2$.

Testing closeness using $O(N^{1/2})$ queries

Accept if $p = q$, reject if $\|p - q\|_1 \geq \epsilon$

Brute force method: estimate $\|p - q\|_1$ with precision $\sim \epsilon$

Testing closeness using $O(N^{1/2})$ queries

Accept if $p = q$, reject if $\|p - q\|_1 \geq \epsilon$

Brute force method: estimate $\|p - q\|_1$ with precision $\sim \epsilon$

Step 1. $\|p - q\|_1 = \sum_{i=1}^{N} |p_i - q_i| = 2\mathbb{E}(x)$

$$x_i = \frac{|p_i - q_i|}{p_i + q_i} \in [0, 1],$$

i is drawn from $(p + q)/2$

Use Monte Carlo method to estimate $\mathbb{E}(x)$

Testing closeness using $O(N^{1/2})$ queries

Accept if $p = q$, reject if $\|p - q\|_1 \geq \epsilon$

Brute force method: estimate $\|p - q\|_1$ with precision $\sim \epsilon$

Step 1. $\|p - q\|_1 = \sum_{i=1}^{N} |p_i - q_i| = 2\mathbb{E}(x)$

$$x_i = \frac{|p_i - q_i|}{p_i + q_i} \in [0, 1],$$

i is drawn from $(p + q)/2$

Use Monte Carlo method to estimate $\mathbb{E}(x)$

Step 2. Show that estimating x_i with precision ϵ requires estimating p_i, q_i with precision $O(\epsilon \max\{p_i, q_i\})$

Testing closeness using $O(N^{1/2})$ queries

Accept if $p = q$, reject if $\|p - q\|_1 \geq \epsilon$

Brute force method: estimate $\|p - q\|_1$ with precision $\sim \epsilon$

Step 1. $\|p - q\|_1 = \sum_{i=1}^{N} |p_i - q_i| = 2\mathbb{E}(x)$

$$x_i = \frac{|p_i - q_i|}{p_i + q_i} \in [0, 1],$$

i is drawn from $(p + q)/2$

Use Monte Carlo method to estimate $\mathbb{E}(x)$

Step 2. Show that estimating x_i with precision ϵ requires estimating p_i, q_i with precision $O(\epsilon \max\{p_i, q_i\})$

Step 3. Use quantum counting to estimate p_i and q_i

Quantum counting [BHMT 00]

Quantum counting [BHMT 00]

Theorem: For any $i \in [N]$ and any precision $\delta > 0$ one can get an estimate $\tilde{p}_i$ which satisfies $|\tilde{p}_i - p_i| \leq \delta$ w.h.p. using

$$M = O(1) \max\left\{\frac{\sqrt{p_i}}{\delta}, \frac{1}{\sqrt{\delta}}\right\}$$

queries to the oracle generating p.

Quantum counting [BHMT 00]

Theorem: For any $i \in [N]$ and any precision $\delta > 0$ one can get an estimate $\tilde{p}_i$ which satisfies $|\tilde{p}_i - p_i| \leq \delta$ w.h.p. using

$$M = O(1)\max\left\{\frac{\sqrt{p_i}}{\delta}, \frac{1}{\sqrt{\delta}}\right\}$$

queries to the oracle generating p.

Step 3. Use quantum counting to estimate p_i and q_i

We need precision $\delta \sim \epsilon \max\{p_i, q_i\}$ which translates into

$$M = \frac{O(1)}{\sqrt{\max\{p_i, q_i\}}} \quad \text{queries}$$

Quantum counting [BHMT 00]

Theorem: For any $i \in [N]$ and any precision $\delta > 0$ one can get an estimate $\tilde{p}_i$ which satisfies $|\tilde{p}_i - p_i| \leq \delta$ w.h.p. using

$$M = O(1) \max \left\{ \frac{\sqrt{p_i}}{\delta}, \frac{1}{\sqrt{\delta}} \right\}$$

queries to the oracle generating p.

Step 3. Use quantum counting to estimate p_i and q_i

We need precision $\delta \sim \epsilon \max\{p_i, q_i\}$ which translates into

$$M = \frac{O(1)}{\sqrt{\max\{p_i, q_i\}}} \quad \text{queries}$$

Step 4. Show that elements with $\max\{p_i, q_i\} \ll 1/N$ are unlikely to appear. Thus $M = O(\sqrt{N})$ queries suffices.

Testing closeness using $O(N^{1/2})$ queries

Accept if $p = q$, reject if $\|p - q\|_1 \geq \epsilon$

Brute force method: estimate $\|p - q\|_1$ with precision $\sim \epsilon$

Step 1. $\|p - q\|_1 = \sum_{i=1}^{N} |p_i - q_i| = 2\mathbb{E}(x)$

$$x_i = \frac{|p_i - q_i|}{p_i + q_i} \in [0, 1],$$

i is drawn from $(p + q)/2$

Use Monte Carlo method to estimate $\mathbb{E}(x)$

Step 2. Show that estimating x_i with precision ϵ requires estimating p_i, q_i with precision $O(\epsilon \max\{p_i, q_i\})$

Step 3. Use quantum counting to estimate p_i and q_i

Quantum counting [BHMT 00]

Theorem: One can estimate $\|p - q\|_1$ with precision ϵ and error probability ω using

$$M = O(1)\,\frac{\sqrt{N}}{\epsilon^4\,\omega^3}$$

queries to the quantum oracles generating p and q

Theorem: One can estimate $\|p - q\|_1$ with precision ϵ and error probability ω using

$$M = O(1)\,\frac{\sqrt{N}}{\epsilon^4\,\omega^3}$$

queries to the quantum oracles generating p and q

Corollary: One can test closeness using $O(\sqrt{N})$ queries.

Theorem: One can estimate $\|p-q\|_1$ with precision ϵ and error probability ω using

$$M = O(1)\,\frac{\sqrt{N}}{\epsilon^4\,\omega^3}$$

queries to the quantum oracles generating p and q

Corollary: One can test closeness using $O(\sqrt{N})$ queries.

Classical lower bounds:

Closeness testing: $\Omega(N^{2/3})$

Estimating $\|p-q\|_1$: $\Omega(N^{1-o(1)})$

Theorem: One can estimate $\|p-q\|_1$ with precision ϵ and error probability ω using

$$M = O(1)\,\frac{\sqrt{N}}{\epsilon^4\,\omega^3}$$

queries to the quantum oracles generating p and q

Corollary: One can test closeness using $O(\sqrt{N})$ queries.

Classical lower bounds:

Closeness testing: $\Omega(N^{2/3})$

Estimating $\|p-q\|_1$: $\Omega(N^{1-o(1)})$

It suggests that the quantum upper bound $O(\sqrt{N})$ for testing closeness might be improved...

Testing uniformity using $O(N^{1/3})$ queries

Accept if $p = I/N$. Reject if $\|p - I/N\|_1 \geq \epsilon$.

Testing uniformity using $O(N^{1/3})$ queries

Accept if $p = I/N$. Reject if $\|p - I/N\|_1 \geq \epsilon$.

What is special about statistics of samples drawn from the uniform distribution?

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

$$r = \sum_{i \in X} p_i \quad \text{— collision probability}$$

Testing uniformity using $O(N^{1/3})$ queries

Accept if $p = I/N$. Reject if $\|p - I/N\|_1 \geq \epsilon$.

What is special about statistics of samples drawn from the uniform distribution?

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

$$r = \sum_{i \in X} p_i \quad \text{— collision probability}$$

p is uniform iff $r \leq \frac{M}{N}$ with probability 1

Testing uniformity using $O(N^{1/3})$ queries

Accept if $p = I/N$. Reject if $\|p - I/N\|_1 \geq \epsilon$.

What is special about statistics of samples drawn from the uniform distribution?

$X = (i_1, \ldots, i_M)$ — a list of M samples drawn from p

$$r = \sum_{i \in X} p_i \quad \text{— collision probability}$$

p is uniform iff $r \leq \frac{M}{N}$ with probability 1

If $M \sim N^{1/3}$ then $r = \frac{M}{N}$ w.h.p.

p is uniform iff $r \leq \frac{M}{N}$ with probability 1

Let's say that p is ϵ-non-uniform iff $\|p - I/N\|_1 \geq \epsilon$

p is uniform iff $r \leq \frac{M}{N}$ with probability 1

Let's say that p is ϵ-non-uniform iff $\|p - I/N\|_1 \geq \epsilon$

Our strategy will be to show that

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad Pr\left[r \geq \frac{M}{N}(1+\delta)\right] \geq \omega$$

for some positive $\delta = \delta(\epsilon)$ and $\omega = \omega(\epsilon)$

p is uniform iff $r \leq \frac{M}{N}$ with probability 1

Let's say that p is ϵ-non-uniform iff $\|p - I/N\|_1 \geq \epsilon$

Our strategy will be to show that

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad \Pr\left[r \geq \frac{M}{N}(1+\delta)\right] \geq \omega$$

for some positive $\delta = \delta(\epsilon)$ and $\omega = \omega(\epsilon)$

Uniformity-Test(M, δ, ω)
Let $X = \{i_1, \ldots, i_M\}$ be a set of M samples from p.
Let $r = \sum_{i \in X} p_i$ be collision probability.
Let $\tilde{r}$ be an estimate of r obtained using the quantum counting algorithm with a relative error δ.
If $\tilde{r} > (1 + \delta)M/N$ then reject. Accept otherwise.

p is uniform iff $r \leq \frac{M}{N}$ with probability 1

Let's say that p is ϵ-non-uniform iff $\|p - I/N\|_1 \geq \epsilon$

Our strategy will be to show that

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad \Pr\left[r \geq \frac{M}{N}(1+\delta)\right] \geq \omega$$

for some positive $\delta = \delta(\epsilon)$ and $\omega = \omega(\epsilon)$

Uniformity-Test(M,δ,ω)
Let $X = \{i_1, \ldots, i_M\}$ be a set of M samples from p.
Let $r = \sum_{i \in X} p_i$ be collision probability.
Let $\tilde{r}$ be an estimate of r obtained using the quantum counting algorithm with a relative error δ.
If $\tilde{r} > (1+\delta)M/N$ then reject. Accept otherwise.

Uniformity-Test(M,δ,ω)
Let $X = \{i_1, \ldots, i_M\}$ be a set of M samples from p.
Let $r = \sum_{i \in X} p_i$ be collision probability.
Let $\tilde{r}$ be an estimate of r obtained using the quantum counting algorithm with a relative error δ.
If $\tilde{r} > (1+\delta)M/N$ then reject. Accept otherwise.

Theorem: Choose parameters of the tester as
$M = 64\epsilon^{-4}N^{1/3}$,
$\delta = \epsilon^2/8$,
$\omega = 1/(2a^a)$ where $a = 64\epsilon^{-4}$.
Then

$$p \text{ is uniform} \quad \Rightarrow \quad \Pr(\text{reject}) \le \omega,$$

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad \Pr(\text{reject}) \ge 3\omega/2$$

Sketch of the proof

We have to prove that

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad \mathrm{P}r\left[r \geq \frac{M}{N}(1+\delta)\right] \geq \omega$$

Sketch of the proof

We have to prove that

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad \Pr\left[r \geq \frac{M}{N}(1+\delta)\right] \geq \omega$$

Simplification 1: we can assume wlog that $p_i \ll N^{-1/3}$. Indeed, if $\exists p_i \sim N^{-1/3}$, such element i will appear in the sample list with a constant probability. Then

$$r \geq p_i \sim N^{-1/3} \gg M/N \sim N^{-2/3}.$$

Sketch of the proof

We have to prove that

$$p \text{ is } \epsilon\text{-non-uniform} \quad \Rightarrow \quad \Pr\left[r \geq \frac{M}{N}(1+\delta)\right] \geq \omega$$

Simplification 1: we can assume wlog that $p_i \ll N^{-1/3}$. Indeed, if $\exists p_i \sim N^{-1/3}$, such element i will appear in the sample list with a constant probability. Then

$$r \geq p_i \sim N^{-1/3} \gg M/N \sim N^{-2/3}.$$

Simplification 2: we can assume wlog that all elements $(i_1, \ldots, i_M)$ in a sample list are distinct. Indeed,

$$\Pr[\exists \alpha \neq \beta : i_\alpha = i_\beta] \leq M^2 \sum_{i=1}^{N} p_i^2 \leq N^{-1/3}.$$

Sketch of the proof

After these simplifications we get

$$r = \sum_{\alpha=1}^{M} p_{i_\alpha}, \quad \text{where } (i_1, \ldots, i_M) \text{ are samples drawn from } p$$

$$\mathbb{E}(r) = M \sum_{i=1}^{N} p_i^2, \quad \mathrm{Var}(r) = M \left(\sum_{i=1}^{N} p_i^3 - [\sum_{i=1}^{N} p_i^2]^2 \right).$$

Sketch of the proof

After these simplifications we get

$$r = \sum_{\alpha=1}^{M} p_{i_\alpha}, \quad \text{where } (i_1, \ldots, i_M) \text{ are samples drawn from } p$$

$$\mathbb{E}(r) = M \sum_{i=1}^{N} p_i^2, \quad \mathrm{Var}(r) = M \left(\sum_{i=1}^{N} p_i^3 - [\sum_{i=1}^{N} p_i^2]^2 \right).$$

Fact 1: p is ϵ-non-uniform $\Rightarrow$ $\mathbb{E}(r) \geq \frac{M}{N}(1+\epsilon^2)$.

Sketch of the proof

After these simplifications we get

$$r = \sum_{\alpha=1}^{M} p_{i_\alpha}, \quad \text{where } (i_1, \ldots, i_M) \text{ are samples drawn from } p$$

$$\mathbb{E}(r) = M \sum_{i=1}^{N} p_i^2, \quad \mathrm{Var}(r) = M \left(\sum_{i=1}^{N} p_i^3 - [\sum_{i=1}^{N} p_i^2]^2 \right).$$

Fact 1: p is ϵ-non-uniform $\quad \Rightarrow \quad \mathbb{E}(r) \geq \frac{M}{N}(1+\epsilon^2)$.

Fact 2: If $\|p\|_\infty \ll N^{-2/3}$ then $\sqrt{\mathrm{Var}(r)} \ll \mathbb{E}(r)$

Sketch of the proof

After these simplifications we get

$$r = \sum_{\alpha=1}^{M} p_{i_\alpha}, \quad \text{where } (i_1, \ldots, i_M) \text{ are samples drawn from } p$$

$$\mathbb{E}(r) = M \sum_{i=1}^{N} p_i^2, \quad \mathrm{Var}(r) = M \left(\sum_{i=1}^{N} p_i^3 - [\sum_{i=1}^{N} p_i^2]^2 \right).$$

Fact 1: p is ϵ-non-uniform $\Rightarrow$ $\mathbb{E}(r) \geq \frac{M}{N}(1+\epsilon^2)$.

Fact 2: If $\|p\|_\infty \ll N^{-2/3}$ then $\sqrt{\mathrm{Var}(r)} \ll \mathbb{E}(r)$

Thus if p has no 'big' elements $p_i \sim N^{-2/3}$ then the standard Chebyshev inequality implies $r \geq (M/N)(1+\delta)$ w.h.p.

Sketch of the proof

Def. An element $i \in [N]$ is called big iff $p_i > 2N^{-2/3}$.

Sketch of the proof

After these simplifications we get

$$r = \sum_{\alpha=1}^{M} p_{i_\alpha}, \quad \text{where } (i_1, \ldots, i_M) \text{ are samples drawn from } p$$

$$\mathbb{E}(r) = M \sum_{i=1}^{N} p_i^2, \quad \mathrm{Var}(r) = M \left(\sum_{i=1}^{N} p_i^3 - [\sum_{i=1}^{N} p_i^2]^2 \right).$$

Fact 1: p is ϵ-non-uniform $\Rightarrow$ $\mathbb{E}(r) \geq \frac{M}{N}(1+\epsilon^2)$.

Fact 2: If $\|p\|_\infty \ll N^{-2/3}$ then $\sqrt{\mathrm{Var}(r)} \ll \mathbb{E}(r)$

Thus if p has no 'big' elements $p_i \sim N^{-2/3}$ then the standard Chebyshev inequality implies $r \geq (M/N)(1+\delta)$ w.h.p.

Sketch of the proof

Def. An element $i \in [N]$ is called big iff $p_i > 2N^{-2/3}$.

Sketch of the proof

After these simplifications we get

$$r = \sum_{\alpha=1}^{M} p_{i_\alpha}, \quad \text{where } (i_1, \ldots, i_M) \text{ are samples drawn from } p$$

$$\mathbb{E}(r) = M \sum_{i=1}^{N} p_i^2, \quad \mathrm{Var}(r) = M \left(\sum_{i=1}^{N} p_i^3 - [\sum_{i=1}^{N} p_i^2]^2 \right).$$

Fact 1: p is ϵ-non-uniform $\quad \Rightarrow \quad \mathbb{E}(r) \geq \frac{M}{N}(1+\epsilon^2)$.

Fact 2: If $\|p\|_\infty \ll N^{-2/3}$ then $\sqrt{\mathrm{Var}(r)} \ll \mathbb{E}(r)$

Thus if p has no 'big' elements $p_i \sim N^{-2/3}$ then the standard Chebyshev inequality implies $r \geq (M/N)(1+\delta)$ w.h.p.

Sketch of the proof

Def. An element $i \in [N]$ is called big iff $p_i > 2N^{-2/3}$.

Sketch of the proof

Def. An element $i \in [N]$ is called big iff $p_i > 2N^{-2/3}$.

$\mathrm{Big} = \{i \in [N] : p_i > 2N^{-2/3}\}$ – a set of big elements

Sketch of the proof

Def. An element $i \in [N]$ is called big iff $p_i > 2N^{-2/3}$.

$\text{Big} = \{i \in [N] : p_i > 2N^{-2/3}\}$ – a set of big elements

$w_{\text{big}} = \sum_{i \in \text{Big}} p_i$ – a probability that i is big

Sketch of the proof

Def. An element $i \in [N]$ is called big iff $p_i > 2N^{-2/3}$.

$\mathrm{Big} = \{i \in [N] : p_i > 2N^{-2/3}\}$ – a set of big elements

$w_{\mathrm{big}} = \sum_{i \in \mathrm{Big}} p_i$ – a probability that i is big

$w_{\mathrm{big}} > N^{-1/3}$ (many big elements)	$w_{\mathrm{big}} \leq N^{-1/3}$ (a few big elements)
n order to make $r > (1+\delta)M/N$ ve need only $O(1)$ big elements n a list of samples $X = (i_1, \ldots, i_M)$. Show that it happens with constant probability (although exp. small in ϵ^{-1}).	Show that a sample $X = (i_1, \ldots, i_M)$ contains no big elements with a constant probability (although exp. small in ϵ^{-1}). Conditioned on having no big elements we already know that $r > (1+\delta)M/N$ w.h.p.

Conclusions

Property	Cl. Upper	Cl. Lower	Q. Upper	Q. Lower
Uniformity	$\tilde{O}(N^{2/3})$	$\Omega(N^{1/2})$	$O(N^{1/3})$	?
Closeness	$\tilde{O}(N^{2/3})$	$\Omega(N^{2/3})$	$O(N^{1/2})$	?
Orthogonality	$O(N^{1/2})$	$\Omega(N^{1/2})$	$O(N^{1/3})$	$\Omega(N^{1/3})$

Open problems:

- Testing closeness: is $O(N^{1/2})$ optimal?
- Quantum lower bounds