# Quantum Communication Complexity of Distribution Testing

ArXiv:2006.14870

François Le Gall

Graduate School of Mathematics
Nagoya University

joint work with Aleksandrs Belovs, Arturo Castellanos, Guillaume Malod
and Alexander Sherstov

SUSTech-Nagoya workshop
22 June 2021

# Classical Distribution Testing

subfield of "classical learning theory"

- ✓ Consider two probability distributions p,q over {1,…,n}
- ✓ We are given access only to a limited number of samples of each distribution
- ✓ Decide if the distributions satisfy some property or <u>are far from</u> satisfying the property

1 sample from p: "i" with probability p(i)
1 sample from q: "j" with probability q(j)

**Closeness testing (l1-norm version)**

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_1 < \varepsilon$ never happens)

$$\| p - q \|_1 = \sum_{i=1}^{n} |p(i) - q(i)|$$

**Closeness testing (l2-norm version)**

Decide if $p = q$ or $\| p - q \|_2 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_2 < \varepsilon$ never happens)

$$\| p - q \|_2 = \sqrt{\sum_{i=1}^{n} |p(i) - q(i)|^2}$$

Today I will mainly consider the case where $\varepsilon$ is a small constant (e.g., $\varepsilon = 1/100$)

# Classical Distribution Testing

subfield of "classical learning theory"

- ✓ Consider two probability distributions p,q over {1,…,n}
- ✓ We are given access only to a limited number of samples of each distribution
- ✓ Decide if the distributions satisfy some property or <u>are far from</u> satisfying the property

1 sample from p: "i" with probability p(i)

1 sample from q: "j" with probability q(j)

**Closeness testing (l1-norm version)**

Decide if p = q or $\| p - q \|_1 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_1 < \varepsilon$ never happens)

→ Sample complexity: $\Theta(n^{2/3})$ (for ε constant)

$\Theta(n^{2/3})$ samples from p and $\Theta(n^{2/3})$ samples from q

Upper bound: [Batu, Fortnow, Rubinfeld, Smith, White 2000]

Lower bound: [Valiant 2008]

Tight bounds for small ε : [Chan, Diakonikolas, Valiant, Valiant 2014] [Diakonikolas and Kane 2016]

**Closeness testing (l2-norm version)**

Decide if p = q or $\| p - q \|_2 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_2 < \varepsilon$ never happens)

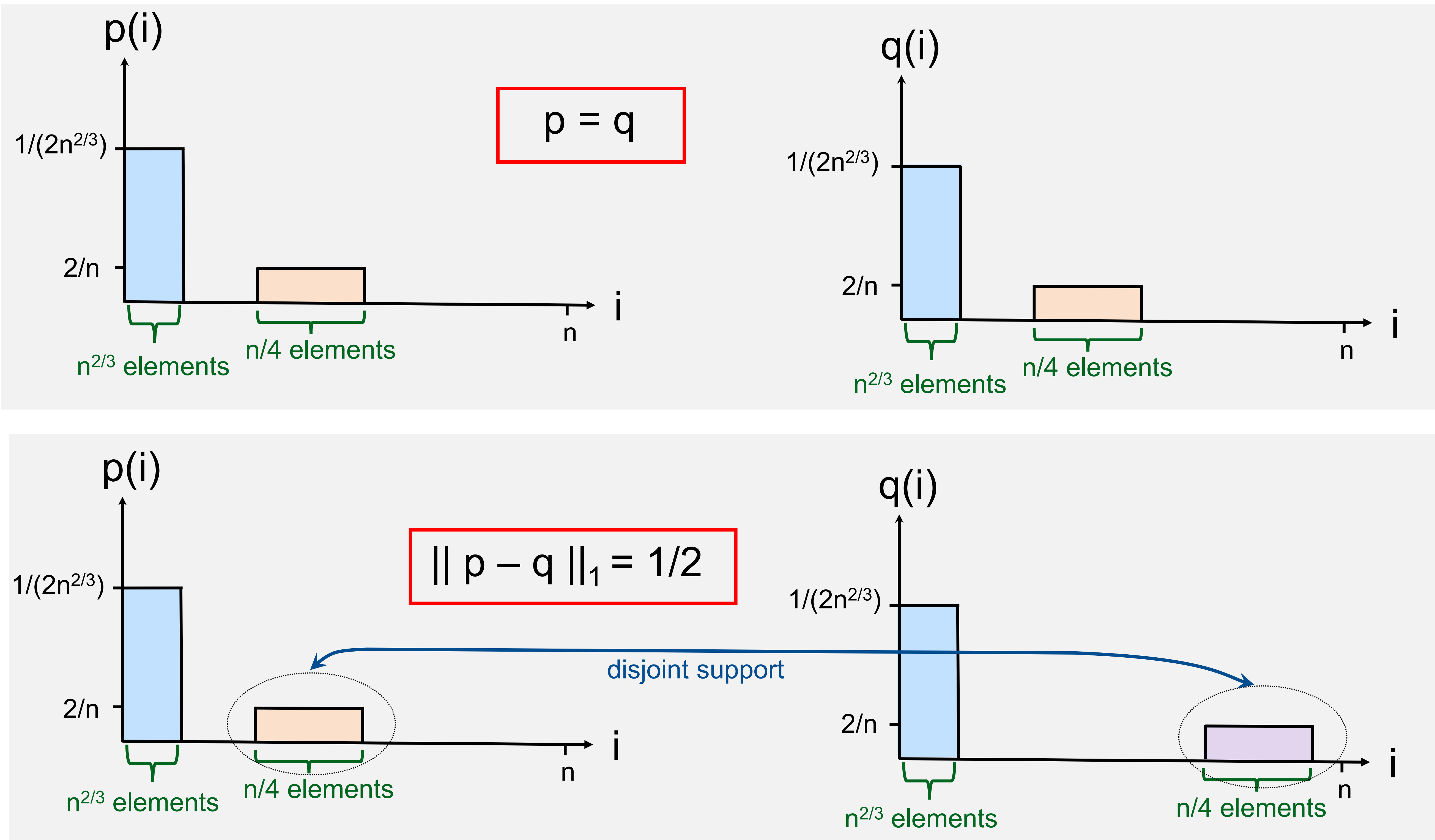
→ Sample complexity: $\Theta(1)$

More precisely: $\Theta(1/\varepsilon^2)$ samples

[Chan, Diakonikolas, Valiant, Valiant 2014]

Today I will mainly consider the case where ε is a small constant (e.g., ε = 1/100)

# Hardest Case l1-Norm Closeness Testing




Claim: distinguishing between the two cases requires $\Theta(n^{2/3})$ samples

# Classical Distribution Testing

subfield of "classical learning theory"

✓ Consider two probability distributions p,q over {1,…,n}

✓ We are given access only to a limited number of samples of each distribution

1 sample from p: "i" with probability p(i)
1 sample from q: "j" with probability q(j)

✓ Decide if the distributions satisfy some property or <u>are far from</u> satisfying the property

**Closeness testing (l1-norm version)**

Decide if p = q or $|| p - q ||_1 \geq \varepsilon$

(Assumption: the case $0 < || p - q ||_1 < \varepsilon$ never happens)

⇒ Sample complexity: $\Theta(n^{2/3})$ (for ε constant)

$\Theta(n^{2/3})$ samples from p and $\Theta(n^{2/3})$ samples from q

Upper bound: [Batu, Fortnow, Rubinfeld, Smith, White 2000]

Lower bound: [Valiant 2008]

Tight bounds for small ε : [Chan, Diakonikolas, Valiant, Valiant 2014] [Diakonikolas and Kane 2016]

**Closeness testing (l2-norm version)**

Decide if p = q or $|| p - q ||_2 \geq \varepsilon$

(Assumption: the case $0 < || p - q ||_2 < \varepsilon$ never happens)

⇒ Sample complexity: $\Theta(1)$

More precisely: $\Theta(1/\varepsilon^2)$ samples

[Chan, Diakonikolas, Valiant, Valiant 2014]

Today I will mainly consider the case where ε is a small constant (e.g., ε = 1/100)

# Quantum Distribution Testing

← subfield of "quantum learning theory"

quantum sample ("purified quantum query-access")

1 quantum sample from p: one copy of the quantum state $\sum_{i=1}^{n} \sqrt{p(i)}\,|i\rangle$

1 quantum sample from q: one copy of the quantum state $\sum_{i=1}^{n} \sqrt{q(i)}\,|i\rangle$

Main criticism: it does not look "fair" to compare classical and quantum learning theories since this concept of quantum sample looks much stronger

**Closeness testing (l1-norm version)**

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_1 < \varepsilon$ never happens)

⇒ Sample complexity: $\Theta(n^{2/3})$ (for ε constant)

Quantum sample complexity:

$\Theta(n^{1/2})$ quantum samples (for ε constant)

[Montanaro 2015], [Gilyen and Li 2020]

**Closeness testing (l2-norm version)**

Decide if $p = q$ or $\| p - q \|_2 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_2 < \varepsilon$ never happens)

⇒ Sample complexity: $\Theta(1)$

More precisely: $\Theta(1/\varepsilon^2)$ samples

Quantum sample complexity:

$\Theta(1/\varepsilon)$ quantum samples

[Montanaro 2015], [Gilyen and Li 2020]

# Quantum Distribution Testing

subfield of "quantum learning theory"

**This work: Quantum Distribution Testing with Classical Samples**

Main criticism: it does not look "fair" to compare classical and quantum learning theories since this concept of quantum sample looks much stronger

**Closeness testing (l1-norm version)**

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_1 < \varepsilon$ never happens)

**Closeness testing (l2-norm version)**

Decide if $p = q$ or $\| p - q \|_2 \geq \varepsilon$

(Assumption: the case $0 < \| p - q \|_2 < \varepsilon$ never happens)

# Classical Communication Complexity of Distribution Testing

**Closeness testing (l1-norm version)**

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

[Andoni, Malkin and Nosatzki 2019] studied the communication complexity of this problem

Alice

input: t classical samples from p

classical communication

⋮

Bob

input: t classical samples from q

How many bits of communication do Alice and Bob need to exchange in order to solve the problem?

- ✓ if $t \approx n^{2/3}$, then we do the same as for the trivial protocol
- ✓ if t is larger (e.g., $t \approx n$), then Alice and Bob can learn from themselves a good approximation of p and q, and then use a protocol specific to these p and q

- ✓ Nothing can be done if $t = o(n^{2/3})$
- ✓ Trivial protocol: Alice sends all its samples to Bob
  - Solves the problem if $t = \Omega(n^{2/3})$
  - Uses $O(t \log n)$ bits of communication (each sample can be encoded by $O(\log n)$ bits)

Theorem [Andoni, Malkin and Nosatzki 2019]

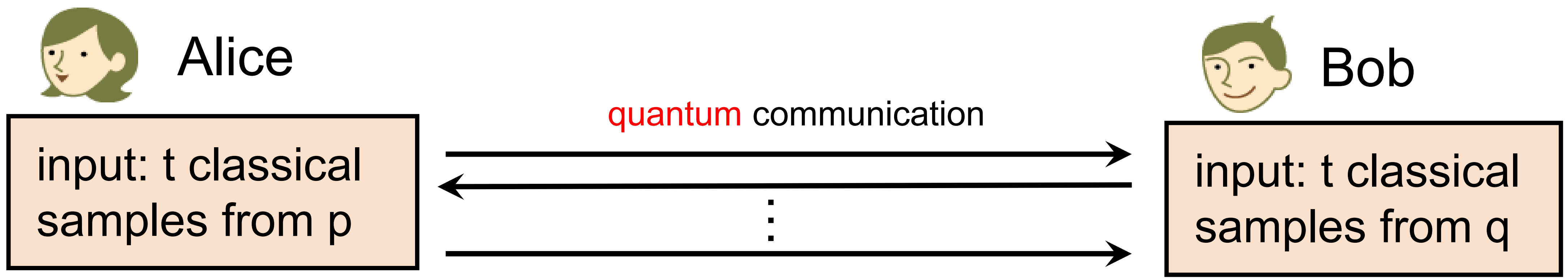
For any $t \in [\Omega(n^{2/3}), n]$, this problem can be solved with high probability using $\tilde{O}((n/t)^2)$ bits of communication. This upper bound is tight.

# Quantum Communication Complexity of Distribution Testing

**Closeness testing (l1-norm version)**

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

[Andoni, Malkin and Nosatzki 2019] studied the communication complexity of this problem




Our result: For any $t \in [\Omega(n^{2/3}), n]$, when $\min(\| p \|_2, \| q \|_2) = O(t/n)$ this problem can be solved with high probability using $\tilde{O}(n/t)$ qubits of communication. This upper bound is tight.

quadradic improvement for low-norm distributions

Theorem [Andoni, Malkin and Nosatzki 2019]: For any $t \in [\Omega(n^{2/3}), n]$, this problem can be solved with high probability using $\tilde{O}((n/t)^2)$ bits of communication. This upper bound is tight.

# Occurrence Vectors

Consider t samples of the distribution $p: \{1,\ldots,n\} \rightarrow [0,1]$

For each $i \in \{1,\ldots,n\}$, let $X_i$ be the number of samples corresponding to element i.

The vector $X = (X_1, X_2, \ldots X_n) \in \{0,1,\ldots,t\}^n$ is called the occurrence vector of these samples.

example: n = 5, samples "1", "3", "1", "2", "5", "3" ⇨ X = (2,1,2,0,1)

Theorem (informal) [Chan, Diakonikolas, Valiant, Valiant 2014]

> Let X denote the occurrence vector of the t samples of p, and let Y denote the occurrence vector of the t samples of q. When $t = \Omega(n^{2/3})$ and $\min(\| p \|_2, \| q \|_2) = O(t/n)$, with high probability a good approximation of $\| X - Y \|_2$ gives a good approximation of $\| p - q \|_2$.

How to use this technique?

> To decide if $\| p - q \|_1 = 0$ or $\| p - q \|_1 \geq \varepsilon$,
>
> decide if $\| p - q \|_2 = 0$ or $\| p - q \|_2 \geq \varepsilon/\sqrt{n}$

# Classical Protocol (for the case min(|| p $||_2$, || q $||_2$) = O(t/n))

Alice

Bob

input: t classical samples from p

classical communication

⋮

input: t classical samples from q

**Classical Protocol for Distribution Closeness Testing from [AMN19]**

1: Fix $\alpha = \Theta(\frac{t\epsilon^2}{n} + 1)$;

2: Alice and Bob each estimate $||p||_2$ and $||q||_2$ up to a factor 2; if the two estimates are not within a factor 4, output "$\epsilon$-FAR";

3: Alice and Bob approximate $\Delta = ||X - Y||_2^2$ up to a $(1+\alpha)$ factor using standard techniques;

4: If $\Delta$ is less than $\tau = \frac{\epsilon^2 t^2}{2n} + 2t$ output "SAME", and otherwise output "$\epsilon$-FAR";

$\tilde{O}(1/\alpha^2) = \tilde{O}((n/t\varepsilon^2)^2)$ bits of communication

Theorem (informal) [Chan, Diakonikolas, Vali... Valiant 2014]

Let X denote the occurrence vector of the t samples of p, and let Y denote the occurrence vector of the t samples of q. When t = $\Omega(n^{2/3})$ and min(|| p $||_2$, || q $||_2$) = $O$(t/n), with high probability a good approximation of || X – Y $||_2$ gives a good approximation of || p – q $||_2$

...holds: for any ...ctly distinguish

If min(|| p $||_2$, || q $||_2$) = Ω(t/n) we do some preprocessing that costs $O((n/t)^2)$ bits of communication

How to use this technique?

Theorem [Andoni, Malkin and Nosatzki 2019]

To decide if || p – q $||_1$ = 0 or || p – q $||_1$ ≥ ε,

decide if || p – q $||_2$ = 0 or || p – q $||_2$ ≥ ε/$\sqrt{n}$

# Quantum Protocol (for the case $\min(\|p\|_2, \|q\|_2) = O(t/n)$)

Alice

Bob

input: t classical samples from p

quantum communication

input: t classical samples from q

**Quantum Protocol for Distribution Closeness Testing**

1: Fix $\alpha = \Theta(\frac{t\epsilon^2}{n} + 1)$;
2: Alice and Bob each estimate $||p||_2$ and $||q||_2$ up to a factor 2; if the two estimates are not within a factor 4, output "$\epsilon$-FAR";
3: Alice and Bob approximate $\Delta = ||X - Y||_2^2$ up to a $(1+\alpha)$ factor using a quantum protocol;
4: If $\Delta$ is less than $\tau = \frac{\epsilon^2 t^2}{2n} + 2t$ output "SAME", and otherwise output "$\epsilon$-FAR";

$\tilde{O}(1/\alpha) = \tilde{O}((n/t\varepsilon^2))$ qubits of communication

~~$O(1/\alpha^2) = O((n/t\varepsilon^2)^2)$ bits of communication~~

**Theorem 3.** [AMN19] *There exists an absolute constant $\gamma_0$ such that the following holds: for any input distributions p and q such that $\min(||p||_2, ||q||_2) \leq \gamma_0 t\epsilon^2/n$, the above protocol correctly distinguish between the case $p = q$ and the case $||p - q||_1 \geq \epsilon$ with probability at least 2/3.*

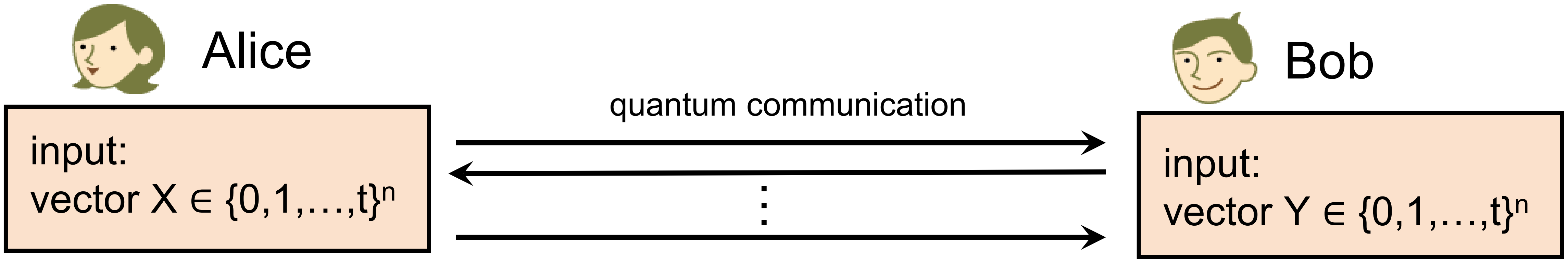
Our result:

For any $t \in [\Omega(n^{2/3}), n]$, when $\min(\|p\|_2, \|q\|_2) = O(t/n)$ this problem can be solved with high probability using $\tilde{O}(n/t)$ qubits of communication. This upper bound is tight.

if $\min(\|p\|_2, \|q\|_2) \gg t/n$ we do some preprocessing that costs $O((n/t)^2)$ bits of communication

too costly!

# Quantum Protocol for (1+α)-Approximation of $\| X - Y \|_2^2$




Goal: for a given precision parameter $\alpha \in [0,1]$, compute a real number d such that

$$(1 - \alpha) \| X - Y \|_2^2 \leq d \leq (1+\alpha) \| X - Y \|_2^2 .$$

Idea: use the AMS technique [Alon, Matias, Szegedy 1999]

Consider a family of $O(n^2)$ functions $h_i$: {1,…,n} → {-1,1} that are 4-wise independent.

for any $(x_1,y_1), (x_2,y_2), (x_3,y_3), (x_4,y_4) \in \{1,\ldots,n\} \times \{-1,1\}$

$$\Pr_i [\, h_i(x_1) = y_1 \wedge h_i(x_2) = y_2 \wedge h_i(x_3) = y_3 \wedge h_i(x_4) = y_4 \,] = 1/16$$

# Quantum Protocol for (1+α)-Approximation of $\| X - Y \|_2^2$

Alice

Bob

input:
vector $X \in \{0,1,\ldots,t\}^n$

quantum communication →
←
⋮
→

input:
vector $Y \in \{0,1,\ldots,t\}^n$

Write $f(i) = \left( \sum_{j=1}^{n} h_i(j) \cdot (X_j - Y_j) \right)^2$

$\mathbb{E}[f(i)] = \mathbb{E}[\underbrace{h_i(1)^2}_{1}(X_1 - Y_1)^2 + \underbrace{h_i(1)h_i(2)}_{\text{0 on average}}(X_1 - Y_1)(X_2 - Y_2) + \cdots]$

Theorem ([Alon, Matias, Szegedy 1999])

If i is taken uniformly at random: $\mathbb{E}[f(i)] = \| X - Y \|_2^2$ and $\mathrm{Var}[f(i)] \le 2 \| X - Y \|_2^4$

Idea: use the AMS technique [Alon, Matias, Szegedy 1999]

Consider a family of $O(n^2)$ functions $h_i$: $\{1,\ldots,n\} \to \{-1,1\}$ that are <u>4-wise independent</u>.

for any $(x_1,y_1), (x_2,y_2), (x_3,y_3), (x_4,y_4) \in \{1,\ldots,n\} \times \{-1,1\}$

$$\Pr_i [\, h_i(x_1) = y_1 \wedge h_i(x_2) = y_2 \wedge h_i(x_3) = y_3 \wedge h_i(x_4) = y_4 \,] = 1/16$$

# Quantum Protocol for (1+α)-Approximation of $\| X - Y \|_2^2$

Alice

input:
vector $X \in \{0,1,\ldots,t\}^n$

quantum communication

⋮

Bob

input:
vector $Y \in \{0,1,\ldots,t\}^n$

Write $f(i) = \left( \sum_{j=1}^{n} h_i(j) \cdot (X_j - Y_j) \right)^2$

Theorem ([Alon, Matias, Szegedy 1999])

If i is taken uniformly at random: $\mathbb{E}[f(i)] = \| X - Y \|_2^2$ and $\mathrm{Var}[f(i)] \le 2 \| X - Y \|_2^4$

Standard techniques

Classically, taking $\Theta(1/\alpha^2)$ values of i and outputting the mean of f(i) gives an (1+α)-approximation of $\| X - Y \|_2^2$ with high probability

Montanaro 2016 (based on quantum amplitude estimation)

There is a quantum algorithm that makes $\Theta(1/\alpha)$ calls to the function f(i) and outputs a (1+α)-approximation of $\| X - Y \|_2^2$ with high probability

1 call to f = O(log n) qubits of communication

# Quantum Protocol (for the case min(|| p $||_2$, || q $||_2$) = O(t/n))

Alice

input: t classical samples from p

quantum communication

⋮

Bob

input: t classical samples from q

**Quantum Protocol for Distribution Closeness Testing**

1: Fix $\alpha = \Theta(\frac{t\epsilon^2}{n} + 1)$;
2: Alice and Bob each estimate $||p||_2$ and $||q||_2$ up to a factor 2; if the two estimates are not within a factor 4, output "$\epsilon$-FAR";
3: Alice and Bob approximate $\Delta = ||X - Y||_2^2$ up to a $(1 + \alpha)$ factor using a quantum protocol;
4: If $\Delta$ is less than $\tau = \frac{\epsilon^2 t^2}{2n} + 2t$ output "SAME", and otherwise output "$\epsilon$-FAR";

$\tilde{O}(1/\alpha) = \tilde{O}((n/t\varepsilon^2))$ qubits of communication

~~$O(1/\alpha^2) = O((n/t\varepsilon^2)^2)$ bits of communication~~

**Theorem 3.** [AMN19] *There exists an absolute constant $\gamma_0$ such that the following holds: for any input distributions p and q such that $\min(||p||_2, ||q||_2) \le \gamma_0 t\epsilon^2/n$, the above protocol correctly distinguish between the case $p = q$ and the case $||p - q||_1 \ge \epsilon$ with probability at least 2/3.*

Our result:

For any t ∈ [$\Omega(n^{2/3})$,n], when min(|| p $||_2$, || q $||_2$) = O(t/n) this problem can be solved with high probability using $\tilde{O}$(n/t) qubits of communication. This upper bound is tight.

# Quantum Communication Complexity of Distribution Testing

**Closeness testing (l1-norm version)**

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

[Andoni, Malkin and Nosatzki 2019] studied the communication complexity of this problem

Alice

input: t classical samples from p

quantum communication

⋮

Bob

input: t classical samples from q

Our result: For any $t \in [\Omega(n^{2/3}), n]$, when $\min(\| p \|_2, \| q \|_2) = O(t/n)$ this problem can be solved with high probability using $\tilde{O}(n/t)$ qubits of communication. This upper bound is tight.

quadradic improvement for low-norm distributions

Theorem [Andoni, Malkin and Nosatzki 2019]: For any $t \in [\Omega(n^{2/3}), n]$, this problem can be solved with high probability using $\tilde{O}((n/t)^2)$ bits of communication. This upper bound is tight.

# Conclusions and Open Problem

Closeness testing (l1-norm version)

Decide if $p = q$ or $\| p - q \|_1 \geq \varepsilon$

in the framework of communication complexity

- ✓ We showed that there exists a quadratic gap between the classical and quantum communication complexity for small norm distributions
- ✓ Our quantum protocol is optimal: we can prove a matching lower bound by a reduction from the gap Hamming distance using a version of the pattern matrix method tailored for partial functions
- ✓ Since all samples are classical samples (only the communication is quantum), this shows a quantum advantage for "quantum learning theory" with classical samples
- ✓ Main question: can we get a quantum advantage when the distributions have large norm?

# Interesting Research Directions

## Secure Protocols

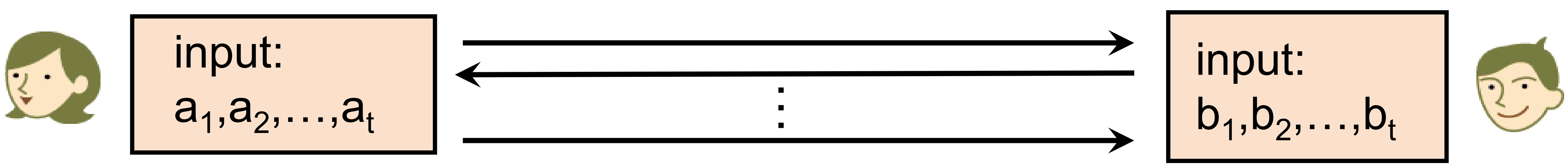
[Andoni, Malkin and Nosatzki 2019] show how to convert their classical protocols into secure protocols. Can we do the same for our quantum protocols?

## Other Properties

[Andoni, Malkin and Nosatzki 2019] also consider Independence Testing. Can we design quantum protocols for this problem as well?

Alice and Bob receive t samples of the distribution p: {1,…,n} x {1,…,n} $\rightarrow$ [0,1]

$(a_1,b_1), \dots, (a_t,b_t)$




Alice and Bob should decide if p is a product distribution of far from any product distribution

What about closeness testing with other norms (e.g., p = q or $|| p - q ||_p \geq \varepsilon$ for $p \in (1,2)$)?

## Quantum Properties?

What is the communication complexity of the following problem: given many copies of a bipartite quantum state ρ, Alice and Bob should decide if ρ is a product state or far from any product state.