# Quantum Probability Oracles & Multidimensional Amplitude Estimation

**Joran van Apeldoorn** ✉
Institute for Information Law, University of Amsterdam, The Netherlands
QuSoft, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

## Abstract

We give a multidimensional version of amplitude estimation. Let $p$ be an $n$-dimensional probability distribution which can be sampled from using a quantum circuit $U_p$. We show that all coordinates of $p$ can be estimated up to error $\varepsilon$ per coordinate using $\widetilde{\mathcal{O}}\left(\frac{1}{\varepsilon}\right)$ applications of $U_p$ and its inverse. This generalizes the normal amplitude estimation algorithm, which solves the problem for $n = 2$. Our results also imply a $\widetilde{\mathcal{O}}(n/\varepsilon)$ query algorithm for $\ell_1$-norm (the total variation distance) estimation and a $\widetilde{\mathcal{O}}\left(\sqrt{n}/\varepsilon\right)$ query algorithm for $\ell_2$-norm. We also show that these results are optimal up to logarithmic factors.

**2012 ACM Subject Classification** Theory of computation → Design and analysis of algorithms; Theory of computation → Quantum computation theory

**Keywords and phrases** quantum algorithms, amplitude estimation, monte carlo

**Digital Object Identifier** 10.4230/LIPIcs.TQC.2021.9

**Funding** *Joran van Apeldoorn*: Supported by the Dutch Research Council (NWO/OCW), as part of the Quantum Software Consortium programme (project number 024.003.037).

**Acknowledgements** I would like to thank Ronald de Wolf, András Gilyén, Ashley Montanaro, and Srinivasan Arunachalam for insightful discussions and comments. I would also like to thank the TQC2021 reviewers for their useful comments.

## 1 Introduction

A central challenge when working with random processes is the estimating of the probability of some event occurring from a bunch of samples. An example from classical computer science is Monte Carlo methods, which try and estimate a value that is hard to compute using a random sampling process. To estimate the probability $p$ of an event occurring using classical samples we can simply sample many times and use the fraction of the outcomes where the event occurred as our estimate. It follows from the Chernoff bound that $\mathcal{O}\left(\frac{\ln(1/\delta)}{\varepsilon^2}\right)$ samples suffice to get an $\varepsilon$ accurate estimate with failure probability at most $\delta$. In fact, it can be shown that this is optimal for classical samples [3].

If we however have access to "quantum samples", that is a unitary that prepares a state that upon measuring would return 1 with probability $p$, than we can improve the number of "samples" needed. The *amplitude estimation* algorithm by Brassard et al. [2] show that $\mathcal{O}\left(\frac{\ln(1/\delta)}{\varepsilon}\right)$ applications of the unitary and it's inverse suffice. This already lays the ground work for numerous general speedups, including for many Monte Carlo methods [9].

Sometimes estimating a single probability is not enough, and we are actually interested in finding a full (discrete) probability distribution. We write $\Delta^n := \{x \in \mathbb{R}^n : x \geq 0 \wedge \|x\|_1 = 1\}$ for the set of all probability distributions on $n$ elements. Let $p \in \Delta^n$, if we take $\mathcal{O}\left(\frac{\ln(n/\delta)}{\varepsilon^2}\right)$ classical samples than for each element $p_i$ we get an estimate $\tilde{p}_i$ such that $|p_i - \tilde{p}_i| \leq \varepsilon$ with error probability at most $\delta/n$. Hence by the union bound over all $i \in n$ it follows that $\|p - \tilde{p}\|_\infty \leq \varepsilon$ with probability at least $1 - \delta$.

16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021).
Editor: Min-Hsiu Hsieh; Article No. 9; pp. 9:1–9:11
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This paper considers the problem of recovering an estimate for a distribution $p \in \Delta^n$ using "quantum samples":

▶ **Definition 1** (Quantum probability oracle). *Let $p \in \Delta^n$ be a probability distribution. We say that $O_p$ is a* quantum probability oracle *for $p$ if*

$$O_p\,|0\rangle = \sum_{i=1}^{n} \sqrt{p_i}\,|i\rangle|\psi_i\rangle$$

*for some quantum states $|\psi_1\rangle, \ldots, |\psi_n\rangle$. That is, applying $O_p$ to the $|0\rangle$ state and measuring the first register is the same as sampling from $p$.*

Throughout the paper we will assume that if we can apply $O_p$, then we can also apply $O_p^{-1}$, and we can do both in a controlled way. Note that this is the case if $O_p$ comes from a randomized classical or quantum algorithm.

We generalize the result of amplitude estimation to $n$-dimensional distributions, showing that an $\varepsilon$-$\ell_\infty$-estimate can be obtained with $\widetilde{\mathcal{O}}\left(\frac{1}{\varepsilon}\right)$ queries to a quantum probability oracle. We do so using a multidimensional version of quantum phase estimation, in a similar manner as the quantum gradient estimation algorithm by Jordan [6, 4]. In fact, we consider estimating the gradient of the function $f(x) = \langle x, p\rangle$.

We also consider $\ell_1$-norm (or total variation distance) and $\ell_2$-norm estimates. We get $\widetilde{\mathcal{O}}\left(\frac{n}{\varepsilon}\right)$ and $\widetilde{\mathcal{O}}\left(\frac{\sqrt{n}}{\varepsilon}\right)$ query algorithms respectively using norm equivalence. In the second part of the paper we give lower bounds that matches the upper bounds up to logarithmic factors for $\ell_1$-norm and $\ell_2$-norm. An $\ell_\infty$-norm lower bound follows from known lower bounds on amplitude estimation. We end the paper with some open questions.

**Table 1** Comparison of known classical sampling bounds and our quantum results for estimating a distribution $p \in \Delta^n$ up to $\varepsilon$ error in a certain norm. Here the $\widetilde{\Theta}\left(\cdots\right)$ hides polylogarithmic factors in $n$ and $1/\varepsilon$. $^\star$The $\ell_2$-norm quantum lower bound only holds when $\varepsilon < 1/(3\sqrt{n})$.

| | Known Classical | Quantum |
|---|---|---|
| $\ell_\infty$ | $\widetilde{\Theta}\left(\frac{1}{\varepsilon^2}\right)$<br>LB:[3] UB:Chernoff | $\widetilde{\Theta}\left(\frac{1}{\varepsilon}\right)$<br>LB: [1][1] UB: Theorem 9 |
| $\ell_2$ | $\widetilde{\Theta}\left(\frac{1}{\varepsilon^2}\right)$<br>LB:[3] UB:[7] | $\widetilde{\mathcal{O}}\left(\min(\frac{\sqrt{n}}{\varepsilon}, \frac{1}{\varepsilon^2})\right)$, $\Omega\left(\frac{\sqrt{n}}{\varepsilon}\right)^\star$<br>LB: Corollary 12 UB: Corollary 10 |
| $\ell_1$ | $\widetilde{\mathcal{O}}\left(\frac{n}{\varepsilon^2}\right)$<br>UB:[7] | $\widetilde{\Theta}\left(\frac{n}{\varepsilon}\right)$<br>LB: Lemma 11 UB: Corollary 10 |

## 2 Upper bound

We show our main result in two steps. First we prove the base result, Theorem 5, which has an almost optimal query complexity but lacks in a few other areas. We then add several improvements to obtain our main result, Theorem 9.

[1] A lower bound on normal amplitude estimation follows from the lower bound on parity given in [1].

## 2.1 Main algorithm

In this section we will show how to obtain an $\varepsilon$-$\ell_\infty$-approximation of $p \in \Delta^n$ using $\mathcal{O}\left(\frac{\ln(n)}{\varepsilon}\right)$ queries to a quantum probability oracle for $p$. To do so we consider the linear function $f : [0,1]^n \to [0,1] : x \mapsto \langle x, p\rangle$. We will show how to construct a specific type of oracle for this function, use known results to convert this to a phase roacle for the function, and then apply multidimensional phase estimation to obtain the gradient $p$.

We will use the following two oracle definitions:

▶ **Definition 2** (Oracles for functions). *Let $f : D \to [0,1]$ be a $[0,1]$ valued function from a discrete domain $D$. A probability oracle for the function $f$ is a unitary $U_f$ that acts as*

$$U_f\,|x\rangle|0\rangle|0\rangle = |x\rangle\left(\sqrt{f(x)}\,|1\rangle|\psi_1^x\rangle + \sqrt{1-f(x)}\,|0\rangle|\psi_0^x\rangle\right).$$

*A phase oracle for the function $f$ is a unitary $U_f$ that acts as*

$$U_f\,|x\rangle = e^{\mathbf{i}f(x)}\,|x\rangle\,.$$

We start by constructing a probability oracle for $f(x) = \langle x, p\rangle$

▶ **Lemma 3.** *Let $U_p$ be a quantum probability oracle for a distribution $p \in \Delta^n$. Let $k \geq 1$ be an integer and let $D = \left\{0, \frac{1}{2^k}, \ldots, \frac{2^k-1}{2^k}\right\}$ be a discretization of $[0,1]$. Then a probability oracle $U_{\tilde{f}}$ for a function $\tilde{f}$ can be constructed such that $\tilde{f}$ is an additive $\mu$-approximation of $f(x) : D^n \to [0,1] : x \mapsto \langle x, p\rangle$ using $2$ queries to $U_p$ and $\widetilde{\mathcal{O}}\left(n\mathrm{polylog}\left(1/\mu\right)\right)$ two-qubit gates. The gate count can be improved to $\mathrm{polylog}\left(n/\mu\right)$ when the input is stored in a QRAM*[2]*.*

**Proof.** We start in a state $|x\rangle|0\rangle|0\rangle|0\rangle|0\rangle$. First we apply $U_f \otimes I$ to obtain

$$|x\rangle\left(\sum_{i=1}^{n}\sqrt{p_i}\,|i\rangle|\psi_i\rangle\right)|0\rangle|0\rangle\,.$$

Now, for each $i \in [n]$ we do the following conditioned on $i$ being in the second register:

1. In the last register, compute an approximation of $2\arcsin\left(\sqrt{x_i}\right)/\pi$ such that the approximation is in $[0,1)$.
2. Conditioned on the first bit of the approximation rotate the second to last register from $|0\rangle$ to $|1\rangle$ over an angle $\pi/4$.
3. Continue for the other bits: rotate over an angle $\pi/8$ conditioned on the second bit, then $\pi/16$, and so on.
4. Uncompute the last register.

Note that we can approximate the arcsin very efficiently, only introducing a logarithmic overhead in terms of the precision. In the end the second to last register will be rotated over an angle very close to $2\arcsin\left(x_i\right)/\pi \times \pi/2 = \arcsin\left(x_i\right)$. We finish the analysis as if the angle was exact. We end up with (after dropping the last register which is now $|0\rangle$ again)

$$|x\rangle\sum_{i=1}^{n}\sqrt{p_i}\,|i\rangle|\psi_i\rangle\left(\sqrt{x_i}\,|1\rangle + \sqrt{1-x_i}\,|0\rangle\right) = |x\rangle\sum_{i=1}^{n}\sqrt{p_i x_i}\,|i\rangle|\psi_i\rangle|1\rangle + \ldots|0\rangle\,.$$

[2] A QRAM allows us to store values in such a way that we can recover them conditioned on an index register using a single QRAM query. While a physical QRAM requires many gates to build, the implementation can likely be highly parallel in a similar manner to classical RAM. When we consider a model with a QRAM we abstract the details of the QRAM away, and count a QRAM query as a single gate, similar to how a classical RAM query is normally counted as a single operation for a classical computer.

The $\ell_2$-norm of the $|1\rangle$ part of te state is $\sqrt{\sum_{i=1}^n \sqrt{p_i x_i}^2} = \sqrt{\langle x,p\rangle}$. We conclude that the state can be written as

$$\sqrt{\langle x,p\rangle}\,|x\rangle|\psi_{x,0}\rangle|0\rangle + \sqrt{1-\langle x,p\rangle}\,|x\rangle|\psi_{x,1}\rangle|1\rangle\,,$$

and hence we have implemented a probability oracle for $f$.

The steps taken for each $i$ can be performed at the same time when $x$ is stored in a QRAM, as this allows us to query $x_i$ in superposition on $|i\rangle$. ◀

For our purposes we will require a phase oracle, not a probability oracle. Luckily, in [4] it was shown that a phase oracle can be constructed from a probability oracle with minimal overhead:

▶ **Lemma 4.** *[4, Corollary 4.1 (Rephrased)] Let $U_f$ be a probability oracle for a function $f: D \to [0,1]$ acting on $q$ qubits. Let $T > 0$. An phase oracle with $\eta$-additive error for $T \cdot f(x)$ can be constructed using $\mathcal{O}(|T| + \log(1/\eta))$ applications of $U_f$ and its inverse, and $\mathcal{O}(q|T| + q\log(1/\eta))$ two-qubit gates.*

We could directly apply quantum gradient calculation [6, 4] now to obtain $p$, but since we have a linear function the result can be obtained using a slightly simpler proof, so we include it for completeness.

▶ **Theorem 5.** *Let $p \in \Delta^n$ and let $U_p$ be a quantum probability oracle acting on $q$ qubits for $p$. Let $\varepsilon > 0$. An approximation $\tilde{p}$ such that $\|p - \tilde{p}\|_\infty \le \varepsilon$ can be found with error probability at most $\delta$ using $\mathcal{O}(\ln(n/\delta)/\varepsilon)$ applications of $U_p$ and $\widetilde{\mathcal{O}}(\ln(\delta)qn/\varepsilon)$ two-qubit gates. The gatecount can be improved to $\widetilde{\mathcal{O}}(\ln(\delta)q(n+1/\varepsilon))$ by using a QRAM.*

**Proof.** Let $k = \lceil \log(4/\varepsilon) \rceil$. Consider the following algorithm:

1. Start in a $n$-register all zero state, where each register as $k$ qubits:

$$|0^k\rangle \dots |0^k\rangle$$

2. Apply Hadamard gates to all qubits to obtain

$$\bigotimes_{i=1}^n \left( \frac{1}{\sqrt{2^k}} \sum_{x_i=0}^{2^k-1} |x_i\rangle \right) = \frac{1}{2^{kn/2}} \sum_{x \in \{0, 2^k-1\}^n} |x\rangle$$

3. Make a phase query for an 1/6-approximation of $f(x) = \langle x,p\rangle$ using Lemma 3 with $\mu < 1/(96\varepsilon)$ and Lemma 4 with $T = 2^k$ and $\eta \le 1/12$ to obtain a state 1/6-close in $\ell_2$-norm to

$$\begin{aligned}\frac{1}{2^{kn/2}} \sum_{x \in \{0,2^k-1\}^n} e^{\mathbf{i}\langle x,p\rangle}\,|x\rangle &= \frac{1}{2^{kn/2}} \sum_{x \in \{0,2^k-1\}^n} e^{\mathbf{i}\sum_i x_i p_i}\,|x\rangle \\ &= \frac{1}{2^{kn/2}} \sum_{x \in \{0,2^k-1\}^n} \left( \prod_{i=1}^n e^{\mathbf{i}x_i p_i} \right) |x\rangle \\ &= \bigotimes_{i=1}^n \left( \frac{1}{\sqrt{2^k}} \sum_{x_i=0}^{2^k-1} e^{\mathbf{i}x_i p_i}\,|x_i\rangle \right)\end{aligned}$$

4. Apply the $k$-qubit inverse QFT to each of the $n$ registers and measure each register.

Note that this algorithm applies $U_p$ a total of $\mathcal{O}\left(\frac{1}{\varepsilon}\right)$ times as per Lemma 4. The gate cost of the phase oracle implementation is $\widetilde{\mathcal{O}}\left(qn/\varepsilon\right)$ (or $\widetilde{\mathcal{O}}\left(q\ln(n)/\varepsilon\right)$ when using QRAM), and the $n$ inverse QFTs require $\mathcal{O}\left(n\ln^2(1/\varepsilon)\right)$ gates[3].

If we ignore the $\ell_2$-error due to the imperfect phase oracle than it would follow from the analysis of phase estimation that we end up with a vector $\tilde{p}$ such that $|p_i - \tilde{p}_i| \leq 4/2^k \leq \varepsilon$ with probability at least $5/6$ per coordinate. Since we incurred at most $1/6$-$\ell_2$-norm error we conclude that $|p_i - \tilde{p}_i| \leq \varepsilon$ with probability at least $2/3$ per coordinate. By repeating $\mathcal{O}\left(\ln(n/\delta)\right)$ times and taking the coordinate wise median, the error probability can be reduced to $\delta/n$. Taking the union bound we get the result from the theorem. ◀

## 2.2 Improvements and tweaks

In this section we give three improvements on the main algorithm. We start by removing the dependence on $n$, leaving only the implicit dependence via $q$ in the gate-complexity. We then show how to get a better query bound when only considering part of a distribution. Finally we show that the algorithm can be tweaked to always return an estimate from $\Delta^n$.

### 2.2.1 Removing the dependence on n

While the main algorithm requires few queries, the time complexity grows linear in $n$. Since the classical algorithm has no dependence linear dependence on $n$ we would hope the same for the quantum algorithm.

The high gate count in the quantum algorithm is due to the fact that we consider all coordinates of $p$, even those with very small or 0 entries. However, to get an $\varepsilon$-$\ell_\infty$-approximation we can ignore all coordinates where the probability is less than $\varepsilon$. This leaves at most $1/\varepsilon$ coordinates to run the algorithm on. To find relevant coordinates we simply use classical samples:

▶ **Lemma 6.** *Let $p \in \Delta^n$, and $\varepsilon, \delta \in (0, 1/3)$. $\mathcal{O}\left(\ln(n/\delta)/\varepsilon\right)$ classical samples suffice to, with error probability at most $\delta$, find all $i \in [n]$ such that $p_i \geq \varepsilon$.*

**Proof.** Consider a single entry $i$ such that $p_i \geq \varepsilon$. After $T$ samples the probability that we have not seen $i$ yet is at most $(1-\varepsilon)^T$. Letting $T = \frac{\ln(\delta\varepsilon)}{\ln(1-\varepsilon)} = \mathcal{O}\left(\frac{\ln(1/(\delta\varepsilon))}{\varepsilon}\right)$ ensures that this error probability is at most $\delta\varepsilon$. Union bounding over the at most $1/\varepsilon$ coordinates gives the result from the lemma. ◀

The lemma shows that the number of coordinates we have to consider in our main algorithm is independent from $n$. As we can simply look at the inner product on those entries, we only get a dependence on $n$ implicitly as $q \geq \log(n)$. In fact, the classical algorithm can be improved using the same method.

### 2.2.2 Learning part of the distribution

Often we will not be interested in all coordinates of $p$, the method from the previous section is an example, but there might be other cases as well. One example is a binary distribution $(p, 1-p)$, where we only need to estimate the first entry. If we know a $p_{\max}$ such that $p \leq p_{\max}$, then amplitude estimation [2] requires $\mathcal{O}\left(\frac{\sqrt{p_{\max}}}{\varepsilon}\right)$ applications of $U_p$.

[3] The square can be removed by approximating the QFT using standard techniques.

Similarly, if we know that $p_i \leq p_{\max}$ for all $i$, then the classical algorithm can be improved by a factor $p_{max}$. Sadly our main algorithm can not be improved by $\sqrt{p_{max}}$ to our knowledge, but we may get a dependence on the sum of the entries in the part of $p$ that we want to estimate.

▶ **Lemma 7.** *Let $p \in \Delta^n$, $\varepsilon \in (0, 1/3)$ and let $S \subseteq [n]$. Let $p_{mt} \geq \sum_{i \in S} p_i$ be the* **m***aximal* **t***otal probability on $S$. We can construct a quantum probability oracle for a distribution $p'' \in \Delta^{n+2}$ using $\mathcal{O}\left(\sqrt{1/p_{mt}}\right)$ applications of $U_p$, membership queries for $S$, and two-qubit gates such that a estimating $p''$ up to $\mathcal{O}\left(\varepsilon/p_{mt}\right)$-$\ell_\infty$-error gives an $\varepsilon$-$\ell_\infty$ error estimate of $p$.*

**Proof.** The main idea is to amplify the probabilities by a factor of $a = \Theta\left(1/p_{mt}\right)$ using $\mathcal{O}\left(\sqrt{a}\right)$ iterations of amplitude amplification. This allows us to take $\varepsilon' = \varepsilon \cdot a$ as a larger error tolerance. However, we need to be careful as we do not know the original $\ell_2$ norm of the "good" part of the state, and hence we do not know the exact amplification that $\mathcal{O}\left(\sqrt{a}\right)$ iterations of amplitude amplification would give, only that it is $\Theta\left(1/p_{mt}\right)$.

We consider a new distribution $p'$ with dimension $n + 2$. The first $n$ coordinates are equal to $p/2$, while the last to coordinates are $p_{mt}/2$ and $(1 - p_{mt})/2$. We can construct a quantum probability oracle $U_{p'}$ for $p'$ using a single controlled application of $U_p$.

Using amplitude amplification we can create an quantum probability oracle $U_{p''}$ for a distribution $p''$ that is equal to $ap'$ on the indices in $S \cup \{n + 1\}$ for some unknown $a \in \Theta\left(1/p_{mt}\right)$. This requires $\mathcal{O}\left(\sqrt{a}\right)$ applications of $U_{p'}$ and membership queries for $S$.

Note that $p''_{n+1} = ap_{mt}/2 = \Theta\left(1\right)$, and in particulair let $L$ be a (constant) lower bound so $p''_{n+1} \geq L$. Now, let $\tilde{p}''$ be a $\frac{\varepsilon L}{8p_{mt}}$-$\ell_\infty$-estimate of $p''$. It follows that $\tilde{p}''_{n+1}$ is an $(1 \pm \frac{\varepsilon}{8p_{mt}})$ multiplicative estimate of $p''_{n+1}$, and hence it gives such a multiplicative estimate $\tilde{a}$ of $a$.

Let $\tilde{p}_i = 2\tilde{p}''_i/\tilde{a}$. We know that $\tilde{p}''_i = p''_i + e_1$ for some error term $e_1$ with $|e_1| \leq L\varepsilon/8p_{mt}$. We also know that $\tilde{a} = a(1 + e_2)$ for some error term $e_2$ with $|e_2| \leq \varepsilon/8p_{mt}$. Hence we know that

$$\begin{aligned}
\tilde{p}_i &= \frac{2\tilde{p}''_i}{\tilde{a}} \\
&= \frac{2(p''_i + e_1)}{a(1 + e_2)} \\
&= \frac{2(\frac{ap_i}{2} + e_1)}{a(1 + e_2)} \\
&= \frac{p_i + 2e_1/a}{(1 + e_2)} \\
&= (p_i + 2e_1/a)(1 + e_3) \\
&= p_i + 2e_1/a + p_ie_3 + 2e_1e_3/a
\end{aligned}$$

where $|e_3| \leq 2|e_2| \leq \varepsilon/4p_{mt}$. We can therefore bound the final error by

$$\begin{aligned}
|2e_1/a + p_ie_3 + 2e_1e_3/a| &\leq |2e_1/a| + |p_ie_3| + |2e_1e_3/a| \\
&\leq 2\frac{L\varepsilon}{4p_{mt}a} + p_{mt}\frac{\varepsilon}{4p_{mt}} + 2\frac{L\varepsilon^2}{32p_{mt}^2a} \\
&\leq 2\frac{L\varepsilon}{8L} + \frac{\varepsilon}{4} + 2\frac{L\varepsilon}{64L} \\
&\leq \varepsilon
\end{aligned}$$

Where we used that $\varepsilon \leq p_{mt}$, as otherwise the problem is trivial, as well as $\frac{1}{p_{mt}a} \leq \frac{1}{2L}$. ◀

### 2.2.3 Returning a probability distribution

Our main algorithm does not always return a $\tilde{p} \in \Delta^n$, all we are promised is that $\|p - \tilde{p}\|_\infty \leq \varepsilon$. The following Lemma shows that we can always convert such a $\tilde{p}$ into a good approximation inside $\Delta^n$.

▶ **Lemma 8.** *Let $p \in \Delta^n$ and let $\tilde{p}$ be such that $\|p - \tilde{p}\|_\infty \leq \varepsilon/8$. Then a $\min(n, 8/\varepsilon)$-sparse $\tilde{p}'$ can be constructed from $\tilde{p}$ such that $\tilde{p}' \in \Delta^n$ and $\|p - \tilde{p}\|_\infty \leq \varepsilon$.*

**Proof.** Let $\tilde{p}'$ be defined by setting all elements in $\tilde{p}$ that are below $\varepsilon/4$ to zero and all elements above 1 to 1, this introduces at most $\varepsilon/4$ extra error in $\ell_\infty$-norm so $\|p - \tilde{p}'\|_\infty \leq \varepsilon/2$.

Now, for an element in $\tilde{p}'$ to be non-zero, the corresponding element of $p$ should be at least $\varepsilon/8$, hence $\tilde{p}'$ has at most $8/\varepsilon$ non-zero elements. Let $k \leq \min(n, 8/\varepsilon)$ be the number of non-zero elements in $\tilde{p}'$. Let $S$ be the sum of the entries in $\tilde{p}'$, so

$$\max(0, 1 - n\varepsilon/2) \leq S \leq 1 + k\varepsilon/4.$$

If $S = 1$ then $\tilde{p}' \in \Delta^n$ so we are done.

If $S > 1$, then we decrease each of the non-zero elements by $(S-1)/k \leq \varepsilon/4$. This introduces at most $\varepsilon/4$ extra error, so the total error is less than $\varepsilon/2 + \varepsilon/4$. Now all elements are non-negative and they sum to 1.

If $S < 1$ and there is an element larger than $1 - \varepsilon/4$, return the distribution that is 1 on the corresponding index and 0 everywhere else. Otherwise we consider two cases, $n \leq 8/\varepsilon$ and $n > 8/\varepsilon$. For the first case, the $\ell_1$-norm error in $\tilde{p}'$ is at most $n\varepsilon/2$, so $1 - S$ is at most $n\varepsilon/2$. Hence, by increasing each coordinate by at most $\varepsilon/2$ we can ensure that the resulting vector is in $\Delta^n$. For the second case we pick $2/\varepsilon$ entries in $\tilde{p}'$, giving preference to the non-zero entries, and increase the picked entries by $\frac{\varepsilon(1-S)}{2} \leq \varepsilon/2$.

Finally, we note that this construction can be implemented in time linear in the input or output sparsity, whichever is larger, times $\log(1/\varepsilon)$. ◀

### 2.2.4 Putting it all together

We can now combine these improvements with our base algorithm to get the following result as a corollary.

▶ **Theorem 9.** *Let $p \in \Delta^n$ and let $U_p$ be a quantum probability oracle acting on $q$ qubits for $p$. Let $\varepsilon > 0$. Let $S \subseteq [n]$ and let $p_{mt}$ be an upperbound on $\sum_{i \in S} p_i$. An $\widetilde{\mathcal{O}}(1/\varepsilon)$-sparse $\tilde{p} \in \Delta^n$ such that $\|p - \tilde{p}\|_\infty \leq \varepsilon$ can be found with error probability at most $\delta > 0$ using $\mathcal{O}\left(\ln(1/\varepsilon\delta)\sqrt{p_{mt}}/\varepsilon\right)$ applications of $U_p$ (and membership queries for $S$) and $\widetilde{\mathcal{O}}\left(q\ln(\delta)\sqrt{p_{mt}}/\varepsilon^2\right)$ two-qubit gates. The gatecount can be improved to $\widetilde{\mathcal{O}}\left(q\ln(\delta)\sqrt{p_{mt}}/\varepsilon\right)$ using QRAM.*

We note that the query complexity matches that of normal amplitude estimation (the query complexity of which is known to be optimal as it can solve the parity problem for a $1/\varepsilon$-bit long string [1]) up to logarithmic factors.

Using the equivalence of norms we can also get upper bounds on the query complexity for $\ell_\rho$ estimates.

▶ **Corollary 10.** *Let $p \in \Delta^n$ and let $U_p$ be a quantum probability oracle acting on $q$ qubits for $p$. Let $\varepsilon > 0$ and $\rho \geq 1$. Let $S \subseteq [n]$ and let $p_{mt}$ be an upperbound on $\sum_{i \in S} p_i$. An $\widetilde{\mathcal{O}}\left(n^{1/\rho}/\varepsilon\right)$-sparse $\tilde{p} \in \Delta^n$ such that $\|p - \tilde{p}\|_\rho \leq \varepsilon$ can be found with error probability at most $\delta > 0$ using $\mathcal{O}\left(\ln(1/\varepsilon\delta)\sqrt{p_{mt}}n^{1/\rho}/\varepsilon\right)$ applications of $U_p$ (and membership queries* *for $S$) and $\widetilde{\mathcal{O}}\left(q\ln(\delta)\sqrt{p_{mt}}n^{2/\rho}/\varepsilon^2\right)$ two-qubit gates. The gatecount can be improved to $\widetilde{\mathcal{O}}\left(q\ln(\delta)\sqrt{p_{mt}}n^{1/\rho}/\varepsilon\right)$ using QRAM.*

We note that this might not always be optimal, in particular in the low-precision regime. For example, classical sampling can produce an $\varepsilon$-$\ell_2$-estimate using $\widetilde{\mathcal{O}}\left(1/\varepsilon^2\right)$ samples as shown by Kamath et al. [7].

## 3 Lower bounds

In this section we will prove lower bounds on the number of applications of $U_p$ that are required to approximate $p$ in different norms. Since the $\ell_\infty$-norm bound follows from known lower bounds on amplitude estimation that can be obtained from the lower bound on parity [1], we focus on the $\ell_1$ and $\ell_2$ norms. We start by proving a lower bound on $\ell_1$-norm estimation.

▶ **Lemma 11.** *Let $\varepsilon \in (0, 1/3)$ and $n \geq 2$. Any algorithm that (with success probability at least $2/3$) for every $p \in \Delta^n$ outputs a $\tilde{p}$ for which $\|p - \tilde{p}\|_1 \leq \varepsilon$ using queries to a quantum probability oracle for $p$, uses at least $\Omega\left(\frac{n}{\varepsilon}\right)$ such queries.*

**Proof.** We assume that $n$ is even as we can always add an extra zero entry. Let $k = \Theta\left(1/\varepsilon\right)$, where $\mu$ will be defined later. Let $x^{(1)}, \cdots, x^{(n/2)} \in \{0,1\}^k$ be such that for all $i$ we have $|x^{(i)}| \in \{k/2, k/2+1\}$. Finding the Hamming weight of a single $x^{(i)}$ solves the majority problem and hence requires $\Omega\left(k\right)$ quantum queries to a standard (binary) oracle for $x^{(i)}$ [1]. We further note that any algorithm that recovers a $n/2$-bit string requires $\Omega\left(n\right)$ quantum queries. Since quantum query complexity is multiplicative under composition [8] it follows that finding all of the $n/2$ Hamming weights requires $\Omega\left(nk\right) = \Omega\left(n/\varepsilon\right)$ quantum queries. Standard techniques can be used to show that finding a constant fraction of the Hamming weights would still require $\Omega\left(n/\varepsilon\right)$ quantum queries, as Grover search can be used to find the “mistakes”.

We now reduce this problem to finding an $\ell_1$-approximation of a probability distribution. Let $p \in \Delta^n$ be given by $p_i = 2\frac{|x^{(i)}|}{nk}$ for $i \leq n/2$ and by $p_i = 2\frac{k-|x^{(i)}|}{nk}$ otherwise. Let $\tilde{p}$ be an $\varepsilon$ approximation of $p$. If $|p_i - \tilde{p}_i| < \frac{1}{nk}$ than we can find $|x^i|$ from $\tilde{p}_i$. As $\tilde{p}$ is an $\varepsilon$-$\ell_1$-norm estimate, it can only be off more than $1/kn = \Theta\left(\varepsilon/n\right)$ on a small constant fraction of the indices, allowing us to find the Hamming weight for all the others.

Finally we show how to implement a quantum probability oracle for $p$. We can sample from $p$ using a classical algorithm as follows:

1. Pick a uniformly random $i \in [n/2]$.
2. Pick a uniformly random $j \in [k]$.
3. If $x_j^{(i)} = 1$ return $i$, if $x_j^{(i)} = 0$ return $i + n/2$.

By replacing the uniformly random picks by the creation of a uniform superposition we get a quantum probability oracle for $p$.

We conclude that $\Omega\left(n/\varepsilon\right)$ queries to a quantum probability oracle for $p$ are required to obtain an $\varepsilon$-$\ell_1$-approximation. ◀

As a corollary we get a lower bound for $\ell_2$-estimates in the high precision regime:

▶ **Corollary 12.** *Let $\varepsilon \in (0, 1/3\sqrt{n})$ and $n \geq 2$. Any algorithm that (with success probability at least $2/3$) for every $p \in \Delta^n$ outputs a $\tilde{p} \in \Delta^n$ for which $\|p - \tilde{p}\|_2 \leq \varepsilon$ using queries to a quantum probability oracle for $p$, uses at least $\Omega\left(\frac{1}{\varepsilon}\right)$ such queries.*

**Proof.** This follows from the fact that $\|p - \tilde{p}\|_1 \leq \sqrt{n}\,\|p - \tilde{p}\|_2$ combined with Lemma 11. ◀

## 4 Open questions

### Estimating the expectation value of stochastic variables

We can identify a stochastic variable over a finite probability distribution $p \in \Delta^n$ with a vector $a \in \mathbb{R}^n$. Here $a_i$ is the value of the stochastic variable on outcome $i$. Hence, the expectation value of the stochastic variable is equal to $\langle a, p\rangle$. If we have $m$ stochastic variables $a^{(1)}, \ldots, a^{(m)}$ then we can write these as the rows of a matrix $A \in \mathbb{R}^{m\times n}$. This leads to the following problem:

> *Let $A \in [-1,1]^{m\times n}$ be a known matrix, let $\varepsilon > 0$ be an error parameter, and let $p \in \Delta^n$ be a unknown probability distribution, accessible via a quantum probability oracle. Output a vector $\tilde{q} \in \mathbb{R}^m$ such that $\|Ap - q\|_\infty \leq \varepsilon$.*

Here we take $A \in [-1,1]^{m\times n}$ for normalization purposes.

Classically this problem can be solved using $\mathcal{O}\left(\frac{\ln(m/\delta)}{\varepsilon^2}\right)$ samples. The argument is similar as before: each expectation value can be estimated with error probability $\delta/m$, and union bounding gives the result. However, our quantum algorithm does not generalize as easily. One way to solve the problem is to apply amplitude estimation $n$ times, but this would use $\widetilde{\mathcal{O}}\left(n\ln(1/\delta)/\varepsilon\right)$ applications of $U_p$. In fact, we can proof the following lower bound:

▶ **Lemma 13.** *Let $\varepsilon \in (0, 1/(3\sqrt{n}))$ and let $n$ be a positive integer power of two. There exists a matrix $A \in \{-1,1\}^{n\times n}$, such that any algorithm that for every $p \in \Delta^n$ (with success probability at least $2/3$) outputs a $\tilde{q} \in \Delta^n$, for which $\|Ap - \tilde{q}\|_\infty \leq \varepsilon$, uses at least $\Omega\left(\frac{\sqrt{n}}{\varepsilon}\right)$ queries to a quantum probability oracle for $p$.*

**Proof.** We let $A \in \{-1,1\}^{n\times n}$ be $\sqrt{n}H^{\otimes \log(n)}$, the rescaled $n$-fold Hadamard, so $\frac{1}{\sqrt{n}}A$ is unitary. Now let $p \in \Delta^n$ be an unknown probability distribution given by a quantum probability oracle. Let $\mathcal{A}$ be an algorithm that uses $T$ queries to a quantum probability oracle for $p$, and outputs an estimate $\tilde{q}$ such that $\|Ap - \tilde{q}\|_\infty \leq \varepsilon$. This $\ell_\infty-$norm estimate also gives an $\ell_2$-norm estimate $\|Ap - \tilde{q}\|_2 \leq \sqrt{n}\varepsilon$. Applying the unitary $\frac{1}{\sqrt{n}}A^T$ gives

$$\frac{1}{\sqrt{n}}\left\|A^TAp - A^T\tilde{q}\right\|_2 \leq \sqrt{n}\varepsilon,$$

and using that $A^TA = nI$ we get

$$\left\|np - A^T\tilde{q}\right\|_2 \leq n\varepsilon.$$

So $\left\|p - \frac{1}{n}A^T\tilde{q}\right\|_2 \leq \varepsilon$, hence from $q$ we can recover an $\varepsilon$-approximation of $p$ in $\ell_2$-norm, which, by Corollary 12 requires at least $\Omega\left(\frac{\sqrt{n}}{\varepsilon}\right)$ queries to a quantum probability oracle for $p$. ◀

We note that the proof, combined with the $\widetilde{\mathcal{O}}\left(\frac{\ln(m/\delta)}{\varepsilon^2}\right)$ classical algorithm for estimating the expectation value of stochastic variables, gives an alternative proof to that of [7] of the fact that $\widetilde{\mathcal{O}}\left(\ln(n/\delta)/\varepsilon^2\right)$ samples suffice for an $\varepsilon$-$\ell_2$-estimate.

Although the lower bound is disappointing, it still leaves open the possibility of an improvement over applying amplitude estimation $n$ times. In particular, when $A = I$ the problem is simply that of $\ell_\infty$-norm estimation, and hence we know that there is an improved algorithm. Slightly more general, if $A$ can be decomposed as $A = RC$ for matrices $R$ and $C$ such that $R$ has a maximal row sum of $r$, and $C$ has a maximal column sum of $c$, then the problem can be solved with $\widetilde{\mathcal{O}}\left(\frac{rc}{\varepsilon}\right)$ queries, by first applying $C/c$ as a leaky Markov chain step, estimating the result in infty norm up to error $\varepsilon/b$, and then applying $R$. It is however unclear for which matrices a good decomposition exists.

### Improvements for partial distributions

While our improved algorithm from Theorem 9 works better when the total probability of seeing a sample we are interested in is low, there is still a discrepancy between the classical dependence on $p_{max}$ and the quantum dependence on $\sqrt{p_{mt}}$.

### Lower bound for low precision $\ell_2$-norm estimates

Our lower bound for $\ell_2$-norm estimates only works for the high precision ($\varepsilon \in \mathcal{O}\left(1/\sqrt{n}\right)$) regime. A $\Omega\left(\frac{1}{\varepsilon}\right)$ lower bound for the $\varepsilon > \frac{1}{\sqrt{n}}$ regime follows from the lower bound on amplitude estimation, but it is an open question whether this may be improved to $\Omega\left(\frac{1}{\varepsilon^2}\right)$.

### Circuit depth

Recent work by Giurgica-Tiron et al. [5] addresses a big disadvantage of amplitude estimation on near term hardware: the circuit depth. While classical probabilities can be estimated by a highly parallel system of logarithmic depth using $\widetilde{\mathcal{O}}\left(1/\varepsilon^2\right)$ processors, quantum amplitude estimation is inherently sequential and takes depth $\widetilde{\mathcal{O}}\left(1/\varepsilon\right)$. Giurgica-Tiron et al. give algorithms that interpolate between these two cases, keeping the depth times the number of oracle queries constant at $\widetilde{\mathcal{O}}\left(1/\varepsilon^2\right)$. It would be interesting to achieve a similar trade-off in the multidimensional case.

### Applications

A natural question is of course that of applications. Since the algorithm works when samples from $p$ are generated by a quantum algorithm, inherently quantum outputs like that of the HHL algorithm, Hamiltonian simulation, or quantum Gibbs sampling might be a good fit. Our new methods allow a lower dependence on the error $\varepsilon$ when performing quantum state tomography on the resulting states than the classical method of simply measuring does.

Another application might lie in distribution learning theory, or more broadly learning theory in general. Here we are given an unknown distribution and are asked to learn certain properties of the distribution. Our estimation algorithm might serve as a new tool to design quantum improvements in this area.

## References

1 R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98. doi:10.1145/502090.502097.

2 Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplification and estimation. *Contemporary Mathematics*, 305, 2002. arXiv:arXiv:quant-ph/0005055.

3 Paul Dagum, Richard Karp, Michael Luby, and Sheldon Ross. An optimal algorithm for monte carlo estimation. *SIAM Journal on Computing*, 29(5):1484–1496, 2000. doi:10.1137/s0097539797315306.

4 András Gilyén, Srinivasan Arunachalam, and Nathan Wiebe. *Optimizing quantum optimization algorithms via faster quantum gradient computation*, pages 1425–1444. SIAM, 2019. doi:10.1137/1.9781611975482.87.

5 Tudor Giurgica-Tiron, Iordanis Kerenidis, Farrokh Labib, Anupam Prakash, and William Zeng. Low depth algorithms for quantum amplitude estimation, 2020. arXiv:2012.03348.

6 Stephen P. Jordan. Fast quantum algorithm for numerical gradient estimation. *Phys. Rev. Lett.*, 95:050501, July 2005. doi:10.1103/PhysRevLett.95.050501.

7 Sudeep Kamath, Alon Orlitsky, Dheeraj Pichapati, and Ananda Theertha Suresh. On learning distributions from their samples. In *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 1066–1100, Paris, France, 2015. PMLR. URL: `http://proceedings.mlr.press/v40/Kamath15.html`.

8 Shelby Kimmel. Quantum adversary (upper) bound. *Chicago Journal of Theoretical Computer Science*, 19(1):1–14, 2013. `doi:10.4086/cjtcs.2013.004`.

9 Ashley Montanaro. Quantum speedup of monte carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015. `doi:10.1098/rspa.2015.0301`.