Quantum Sub-Gaussian Mean Estimator

Yassine Hamoudi
Université de Paris, IRIF, CNRS, F-75013 Paris, France

Abstract

We present a new quantum algorithm for estimating the mean of a real-valued random variable obtained as the output of a quantum computation. Our estimator achieves a nearly-optimal quadratic speedup over the number of classical i.i.d. samples needed to estimate the mean of a heavy-tailed distribution with a sub-Gaussian error rate. This result subsumes (up to logarithmic factors) earlier works on the mean estimation problem that were not optimal for heavy-tailed distributions [9, 8], or that require prior information on the variance [23, 32, 22]. As an application, we obtain new quantum algorithms for the (ϵ, δ)-approximation problem with an optimal dependence on the coefficient of variation of the input random variable.

2012 ACM Subject Classification Theory of computation → Quantum computation theory

Keywords and phrases Quantum algorithm, statistical analysis, mean estimator, sub-Gaussian estimator, (ϵ, δ)-approximation, lower bound

Digital Object Identifier 10.4230/LIPIcs.ESA.2021.50

Funding This research was supported in part by the ERA-NET Cofund in Quantum Technologies project QuantAlgo and the French ANR Blanc project RDAM.

Acknowledgements The authors want to thank the anonymous referees for their valuable comments and suggestions which helped to improve this paper.

1 Introduction

The problem of estimating the mean μ of a real-valued random variable X given *i.i.d.* samples from it is one of the most basic tasks in statistics and in the Monte Carlo method. The properties of the various classical *mean estimators* are well understood. The standard non-asymptotic criterion used to assess the quality of an estimator is formulated as the following *high probability deviation bound*: upon performing n random experiments that return n samples from X, and given a failure probability $\delta \in (0, 1)$, what is the smallest error $\epsilon(n, \delta, X)$ such that the output $\widetilde{\mu}$ of the estimator satisfies $|\widetilde{\mu} - \mu| > \epsilon(n, \delta, X)$ with probability at most δ? Under the standard assumption that the unknown random variable X has a finite variance σ^2, the best possible performances are obtained by the so-called *sub-Gaussian estimators* [30] that achieve the following deviation bound

$$\Pr\left[|\widetilde{\mu} - \mu| > L\sqrt{\frac{\sigma^2 \log(1/\delta)}{n}}\right] \leq \delta \tag{1}$$

for some constant L. The term "sub-Gaussian" reflects that these estimators have a Gaussian tail even for non-Gaussian distributions. The most well-known sub-Gaussian estimator is arguably the *median-of-means* [35, 27, 2], which consists of partitioning the n samples into roughly $\log(1/\delta)$ groups of equal size, computing the empirical mean over each group, and returning the median of the obtained means.

The process of generating a random sample from X is generalized in the quantum model by assuming the existence of a unitary operator U where $U|0\rangle$ coherently encodes the distribution of X. A *quantum experiment* is then defined as one application of this operator or its inverse. The celebrated quantum amplitude estimation algorithm [9] provides a way to estimate the mean of any *Bernoulli* random variable by performing fewer experiments

29th Annual European Symposium on Algorithms (ESA 2021).
Editors: Petra Mutzel, Rasmus Pagh, and Grzegorz Herman; Article No. 50; pp. 50:1–50:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

than with any classical estimator. Yet, for general distributions, the existing quantum mean estimators either require additional information on the variance [23, 32, 22] or are less performant than the classical sub-Gaussian estimators when the distribution is heavy tailed [9, 38, 8, 32]. These results leave open the existence of a general quantum speedup for the mean estimation problem. We address this question by introducing the concept of *quantum sub-Gaussian estimators*, defined through the following deviation bound

$$\Pr\left[|\widetilde{\mu} - \mu| > L\frac{\sigma \log(1/\delta)}{n}\right] \leq \delta \tag{2}$$

for some constant L. We give the first construction of a quantum estimator that achieves this bound up to a logarithmic factor in n. Additionally, we prove that it is impossible to go below that deviation level. This result provides a clear equivalent of the concept of sub-Gaussian estimator in the quantum setting.

A second important family of mean estimators addresses the (ϵ, δ)-*approximation* problem, where given a fixed relative error $\epsilon \in (0, 1)$ and a failure probability $\delta \in (0, 1)$ the goal is to output a mean estimate $\widetilde{\mu}$ such that

$$\Pr[|\widetilde{\mu} - \mu| > \epsilon|\mu|] \leq \delta. \tag{3}$$

The aforementioned sub-Gaussian estimators do not quite answer this question since the number of experiments they require (respectively $n = \Omega\big((\frac{\sigma}{\epsilon\mu})^2 \log(1/\delta)\big)$ and $n = \widetilde{\Omega}\big(\frac{\sigma}{\epsilon|\mu|} \log(1/\delta)\big)\big)$ depends on the *unknown* quantities σ and μ. Sometimes a good upper bound is known on the *coefficient of variation* $|\sigma/\mu|$ and can be used to parametrize a sub-Gaussian estimator. Otherwise, the standard approach is based on *sequential analysis* techniques, where the number of experiments is chosen adaptively depending on the results of previous computations. Given a random variable distributed in $[0, 1]$, the optimal classical estimators perform $\Theta\big(\big((\frac{\sigma}{\epsilon\mu})^2 + \frac{1}{\epsilon\mu}\big) \log(1/\delta)\big)$ random experiments *in expectation* [17] for computing an (ϵ, δ)-approximation of μ. We construct a quantum estimator that reduces this number to $\widetilde{\Theta}\big(\big(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\big) \log(1/\delta)\big)$ and we prove that it is optimal.

1.1 Related work

There is an extensive literature on classical sub-Gaussian estimators and we refer the reader to [30, 15, 12, 18, 28] for an overview of the main results and recent improvements. We point out that the *empirical mean* estimator is not sub-Gaussian, although it is optimal for Gaussian random variables [37, 15]. The non-asymptotic performances of the empirical mean estimator are captured by several standard concentration bounds such as the Chebyshev, Chernoff and Bernstein inequalities.

There is a series of quantum mean estimators [21, 1, 8] that get close to the bound $\Pr\Big[|\widetilde{\mu} - \mu| > L\frac{\log(1/\delta)}{n}\Big] \leq \delta$ for any random variable distributed in $[0, 1]$ and some constant L. Similar results hold for numerical integration problems [1, 36, 23, 39, 24]. The amplitude estimation algorithm [9, 38] leads to a sharper bound of $\Pr\Big[|\widetilde{\mu} - \mu| > L\Big(\frac{\sqrt{\mu(1-\mu)}\log(1/\delta)}{n} + \frac{\log(1/\delta)^2}{n^2}\Big)\Big] \leq \delta$ (see Proposition 12) when X is distributed in $[0, 1]$. Nevertheless, the quantity $\mu(1 - \mu)$ is always larger than or equal to the variance σ^2. The question of improving the dependence on σ^2 was considered in [23, 32, 22]. The estimators of [23, 32] require to know an upper bound Σ on the standard deviation σ, whereas [22] needs an upper bound Δ on the coefficient of variation σ/μ (for non-negative random variables). The performances of these estimators are captured (up to logarithmic factors) by the deviation bound given in Equation (2) with σ replaced by Σ and $\mu\Delta$ respectively.

The (ϵ, δ)-approximation problem has been addressed by several classical works such as [17, 31, 20, 26]. In the quantum setting, there is a variant [9, Theorem 15] of the amplitude estimation algorithm that performs $O(\log(1/\delta)/(\epsilon\sqrt{\mu}))$ experiments in expectation to compute an (ϵ, δ)-approximate of the mean of a random variable distributed in $[0, 1]$ (see Theorem 7 and Proposition 15). However, the complexity of this estimator does not scale with σ. Given an upper bound Δ on σ/μ, the estimator of [22] can be used to compute an (ϵ, δ)-approximate with roughly $\widetilde{O}(\Delta \log(1/\delta)/\epsilon)$ quantum experiments if the random variable is non-negative.

We note that the related problem of estimating the mean with *additive* error ϵ, that is $\Pr[|\widetilde{\mu} - \mu| > \epsilon] \leq \delta$, has also been considered by several authors. The optimal number of experiments is $\Theta(\log(1/\delta)/\epsilon^2)$ classically [14] and $\Theta(1/\epsilon)$ quantumly [34] (with failure probability $\delta = 1/3$). These bounds do not depend on unknown parameters (as opposed to the relative error case), thus sequential analysis techniques are unnecessary here. Montanaro [32] also described an estimator that performs $\widetilde{O}(\Sigma \log(1/\delta)/\epsilon)$ quantum experiments given an upper bound Σ on the standard deviation σ.

1.2 Contributions and organization

We first formally define the input model in Section 2.1. We introduce the concept of "q-random variable" (Definition 3) to describe a random variable that corresponds to the output of a quantum computation. We measure the complexity of an algorithm by counting the number of *quantum experiments* (Definition 4) it performs with respect to a q-random variable. We also introduce some needed tools in Section 2.2. Next, we construct a quantum algorithm for estimating the quantiles of a q-random variable in Section 3, and we use it in Section 4 to design the following quantum sub-Gaussian estimator.

Theorem 13 (Restated). *There exists a quantum algorithm with the following properties. Let X be a q-random variable with mean μ and variance σ^2, and set as input a time parameter n and a real $\delta \in (0, 1)$ such that $n \geq \log(1/\delta)$. Then, the algorithm outputs a mean estimate $\widetilde{\mu}$ such that $\Pr\left[|\widetilde{\mu} - \mu| > \frac{\sigma \log(1/\delta)}{n}\right] \leq \delta$ and it performs $O(n \log^{3/2}(n) \log\log(n))$ quantum experiments.*

Then we turn our attention to the (ϵ, δ)-approximation problem in Section 5. In case we have an upper bound Δ on the coefficient of variation $|\sigma/\mu|$, we directly use our sub-Gaussian estimator to obtain an algorithm that performs $\widetilde{O}\big(\frac{\Delta}{\epsilon} \log(1/\delta)\big)$ quantum experiments (Corollary 14). Next, we consider the more subtle parameter-free setting where there is no prior information about the input random variable, except that it is distributed in $[0, 1]$. In this case, the number of experiments is chosen *adaptively*, and the bound we get is stated in expectation.

Theorem 16 (Restated). *There exists a quantum algorithm with the following properties. Let X be a q-random variable distributed in $[0, 1]$ with mean μ and variance σ^2, and set as input two reals $\epsilon, \delta \in (0, 1)$. Then, the algorithm outputs a mean estimate $\widetilde{\mu}$ such that $\Pr[|\widetilde{\mu} - \mu| > \epsilon\mu] \leq \delta$, and it performs $\widetilde{O}\big(\big(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\big) \log(1/\delta)\big)$ quantum experiments in expectation.*

Finally, we prove several lower bounds in Section 6 that match the complexity of the above estimators. We also consider the weaker input model where one is given copies of a quantum state encoding the distribution of X. We prove that no quantum speedup is achievable in this setting (Theorem 22).

1.3 Proof overview

Sub-Gaussian estimator. Our approach (Theorem 13) combines several ideas used in previous classical and quantum mean estimators. In this section, we simplify the exposition by assuming that the random variable X is non-negative and by replacing the variance σ^2 with the second moment $\mathbb{E}[X^2]$. We also take the failure probability δ to be a small constant. Our starting point is a variant of the *truncated mean estimators* [6, 12, 30]. Truncation is a process that consists of replacing the samples larger than some threshold value with a smaller number. This has the effect of reducing the tail of the distribution, but also of changing its expectation. Here we study the effect of replacing the values larger than some threshold b with 0, which corresponds to the new random variable $Y = X\mathbb{1}_{X \leq b}$. We consider the following classical sub-Gaussian estimator that we were not able to find in the literature: set $b = \sqrt{n\mathbb{E}[X^2]}$ and compute the empirical mean of n samples from Y. By a simple calculation, one can prove that the expectation of the removed part is at most $\mathbb{E}[X - Y] \leq \mathbb{E}[X^2]/b = \sqrt{\mathbb{E}[X^2]/n}$. Moreover, using Bernstein's inequality and the boundedness of Y, the error between the output estimate and $\mathbb{E}[Y]$ is on the order of $\sqrt{\mathbb{E}[X^2]/n}$. These two facts together imply that the overall error for estimating $\mathbb{E}[X]$ is indeed of a sub-Gaussian type. This approach can be carried out in the quantum model by performing the truncation in superposition. This is similar to what is done in previous quantum mean estimators [23, 32, 22]. In order to obtain a quantum speedup, one must balance the truncation level differently by taking $b = n\sqrt{\mathbb{E}[X^2]}$. Then, by a clever use of amplitude estimation discovered by Heinrich [23], the expectation of Y can be estimated with an error on the order of $\sqrt{\mathbb{E}[X^2]}/n$. The main drawback of this estimator is that it requires the knowledge of $\mathbb{E}[X^2]$ to perform the truncation. In previous work [23, 32, 22], the authors made further assumptions on the variance to be able to approximate b. Here, we overcome this issue by choosing the truncation level b differently. Borrowing ideas from classical estimators [30], we define b as the quantile value that satisfies $\Pr[X \geq b] = 1/n^2$. This quantile is always smaller than the previous threshold value $n\sqrt{\mathbb{E}[X^2]}$. Moreover, it can be shown that the removed part $\mathbb{E}[X - Y]$ is still on the order of $\sqrt{\mathbb{E}[X^2]}/n$. We give a new quantum algorithm for approximating this quantile with roughly n quantum experiments (Theorem 11), whereas it would require n^2 random experiments classically. Our quantile estimation algorithm builds upon the quantum minimum finding algorithm of Dürr and Høyer [19, 3] and the kth-smallest element finding algorithm of Nayak and Wu [34]. Importantly, it does not require any knowledge about $\mathbb{E}[X^2]$.

(ϵ, δ)-Approximation without side information. We follow an approach similar to that of a classical estimator described in [17]. Our algorithm (Theorem 16) uses the quantum sub-Gaussian estimator and the quantum *sequential Bernoulli estimator* described in Proposition 15. The latter estimator can estimate the mean μ of a random variable X distributed in $[0, 1]$ with constant relative error by performing $O(1/\sqrt{\mu})$ quantum experiments in expectation. The first step of the (ϵ, δ)-approximation algorithm is to compute a rough estimate $\widehat{\mu}$ of μ with the sequential Bernoulli estimator. Then, the variance σ^2 of X is estimated by using again the sequential Bernoulli estimator on the random variable $(X - X')/2$ (where X' is an independent copy of X). The latter estimation is stopped if it uses more than $O(1/\sqrt{\epsilon\widehat{\mu}})$ quantum experiments. We show that if $\sigma^2 \geq \Omega(\epsilon\mu)$ then the computation is not stopped and the resulting estimate $\widetilde{\sigma}^2$ is close to σ^2 with high probability. Otherwise, it is stopped with high probability and we set $\widetilde{\sigma} = 0$. Finally, the quantum sub-Gaussian estimator is used with the parameter $n \approx \max\left(\frac{\widetilde{\sigma}}{\epsilon\widehat{\mu}}, \frac{1}{\sqrt{\epsilon\widehat{\mu}}}\right)$ to obtain a refined estimate $\widetilde{\mu}$ of μ. The choice of the first (resp. second) term in the maximum value implies that $|\widetilde{\mu} - \mu| \leq \epsilon\mu$ with high probability when the variance σ^2 is larger (resp. smaller) than $\epsilon\mu$. In order to upper bound the expected number of experiments performed by this estimator, we show in Proposition 15 that the estimates $\widehat{\mu}$ and $\widetilde{\sigma}$ obtained with the sequential Bernoulli estimator satisfy the expectation bounds $\mathbb{E}[1/\widehat{\mu}] \leq 1/\mu$, $\mathbb{E}[\widetilde{\sigma}] \leq \sigma$ and $\mathbb{E}[1/\sqrt{\widehat{\mu}}] \leq 1/\sqrt{\mu}$.

Lower bounds. We sketch the proof of optimality of the quantum sub-Gaussian estimator (Theorem 18). The lower bound is proved in the stronger quantum query model, which allows us to extend it to all the other models mentioned in Section 2.1. Our approach is inspired by the truncation level chosen in the algorithm. Given σ and n, we consider the two distributions p_0 and p_1 that output respectively $\frac{n\sigma}{\sqrt{1-1/n^2}}$ and $\frac{-n\sigma}{\sqrt{1-1/n^2}}$ with probability $1/n^2$, and 0 otherwise. The two distributions have variance σ^2 and the distance between their means is larger than $\frac{2\sigma}{n}$. Thus, any estimator that satisfies the bound $\Pr\left[|\widetilde{\mu} - \mu| > \frac{\sigma}{n}\right] \leq \frac{1}{3}$ can distinguish between p_0 and p_1 with constant success probability. However, we show by a reduction to Quantum Search that it requires at least $\Omega(n)$ quantum experiments to distinguish between two distributions that differ with probability at most $1/n^2$.

2 Preliminaries

2.1 Model of input

The input to the mean estimation problem is represented by a real-valued random variable X defined on some probability space. A classical estimator accesses this input by obtaining n *i.i.d* samples of X. In this section, we describe the access model for quantum estimators and we compare it to previous models suggested in the literature. We only consider finite probability spaces for finite encoding reasons. First, we recall the definition of a random variable, and we define a classical model of access called a *random experiment.*

▶ **Definition 1** (Random variable)**.** *A* finite random variable *is a function $X : \Omega \to E$ for some probability space (Ω, p), where Ω is a finite sample set, $p : \Omega \to [0, 1]$ is a probability mass function and $E \subset \mathbb{R}$ is the support of X. As is customary, we will often omit to mention (Ω, p) when referring to the random variable X.*

▶ **Definition 2** (Random experiment)**.** *Given a random variable X on a probability space (Ω, p), we define a* random experiment *as the process of drawing a sample $\omega \in \Omega$ according to p and observing the value of $X(\omega)$.*

We now introduce the concept of "q-random variable" to represent a quantum process that outputs a real number.

▶ **Definition 3** (q-random variable)**.** *A* q-variable *is a triple $(\mathcal{H}, U, M)$ where $\mathcal{H}$ is a finite-dimensional Hilbert space, U is a unitary transformation on $\mathcal{H}$, and $M = \{M_x\}_{x \in E}$ is a projective measurement on $\mathcal{H}$ indexed by a finite set $E \subset \mathbb{R}$. Given a random variable X on a probability space (Ω, p), we say that a q-variable $(\mathcal{H}, U, M)$* generates *X when,*

1. *$\mathcal{H}$ is a finite-dimensional Hilbert space with some basis $\{|\omega\rangle\}_{\omega \in \Omega}$ indexed by Ω.*
2. *U is a unitary transformation on $\mathcal{H}$ such that $U|0\rangle = \sum_{\omega \in \Omega} \sqrt{p(\omega)}|\omega\rangle$.*
3. *$M = \{M_x\}_x$ is the projective measurement on $\mathcal{H}$ defined by $M_x = \sum_{\omega : X(\omega) = x} |\omega\rangle\langle\omega|$.*

A random variable X is a q-random variable *if it is generated by some q-variable $(\mathcal{H}, U, M)$.*

We stress that the sample space Ω may not be known explicitly, and we do not assume that it is easy to perform a measurement in the $\{|\omega\rangle\}_{\omega \in \Omega}$ basis for instance. Often, we are given a unitary U such that $U|0\rangle = \sum_{x \in E} \sqrt{p(x)}|\psi_x\rangle|x\rangle$ for some unknown garbage unit state $|\psi_x\rangle$, together with the measurement $M = \{I \otimes |x\rangle\langle x|\}_{x\in E}$. In this case, we can consider the q-random variable X defined on the probability space (Ω, p) where $\Omega = \{|\psi_x\rangle|x\rangle\}_{x\in E}$ and $X(|\psi_x\rangle|x\rangle) = x$.

We further assume that there exist two quantum oracles, defined below, for obtaining information on the function $X : \Omega \to E$. These two oracles can be efficiently implemented if we have access to a quantum evaluation oracle $|\omega\rangle|0\rangle \mapsto |\omega\rangle|X(\omega)\rangle$ for instance. The rotation oracle (Assumption 2) has been extensively used in previous quantum mean estimators [38, 8, 32, 22]. The comparison oracle (Assumption 1) is needed in our work to implement the quantile estimation algorithm.

▶ **Assumption 1** (Comparison oracle). Given a q-random variable X on a probability space (Ω, p), and any two values $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$ such that $a < b$, there is a unitary operator $C_{a,b}$ acting on $\mathcal{H} \otimes \mathbb{C}^2$ such that for all $\omega \in \Omega$,

$$C_{a,b}(|\omega\rangle|0\rangle) = \begin{cases} |\omega\rangle|1\rangle & \text{when } a < X(\omega) \leq b, \\ |\omega\rangle|0\rangle & \text{otherwise.} \end{cases}$$

▶ **Assumption 2** (Rotation oracle). Given a q-random variable X on a probability space (Ω, p), and any two values $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$ such that $a < b$, there is a unitary operator $R_{a,b}$ acting on $\mathcal{H} \otimes \mathbb{C}^2$ such that for all $\omega \in \Omega$,

$$R_{a,b}(|\omega\rangle|0\rangle) = \begin{cases} |\omega\rangle\left(\sqrt{1 - \left|\frac{X(\omega)}{b}\right|}|0\rangle + \sqrt{\left|\frac{X(\omega)}{b}\right|}|1\rangle\right) & \text{when } a < X(\omega) \leq b, \\ |\omega\rangle|0\rangle & \text{otherwise.} \end{cases}$$

We now define the measure of complexity used to count the number of accesses to a q-random variable, which are referred to as *quantum experiments*.

▶ **Definition 4** (Quantum experiment). *Let X be a q-random variable that satisfies Assumptions 1 and 2. Let $(\mathcal{H}, U, M)$ be a q-variable that generates X. We define a* quantum experiment *as the process of applying any of the unitaries U, $C_{a,b}$, $R_{a,b}$ (for any values of $a < b$), their inverses or their controlled versions, or performing a measurement according to M.*

Note that a random experiment (Definition 2) can be simulated with two quantum experiments by computing the state $U|0\rangle$ and measuring it according to M. We briefly mention two other possible input models. First, some authors [21, 34, 23, 10, 16, 8, 29] consider the stronger query model where p is the uniform distribution and a quantum evaluation oracle is provided for the function $\omega \mapsto X(\omega)$. A second model tackles the problem of *learning from quantum states* [11, 5, 4], where the input consists of several copies of $\sum_{x\in E} \sqrt{\Pr[X = x]}|x\rangle$ (we do not have access to a unitary preparing that state). We show in Theorem 22 that no quantum speedup is achievable for our problem in the latter setting.

2.2 Tools

We will use a variant of the amplitude amplification algorithm that does not need a time parameter n as input. We call it the "sequential amplitude amplification" algorithm in reference to *sequential analysis*. The original version of this algorithm was analysed in Theorem 3 of [7, 9], with a bound on the expected complexity $\mathbb{E}[T]$. We propose a slightly different version that allows us to bound $\mathbb{E}[T^2]$ and $\mathbb{E}[1/T]$ (note that $\mathbb{E}[T] \leq \sqrt{\mathbb{E}[T^2]}$). The algorithm and its analysis are deferred to the extended version of this paper.

▶ **Theorem 5** (Sequential amplitude amplification). *Let U be a unitary quantum algorithm and let Π be a projection operator. Define the number $p \in [0,1]$ and the two unit states $|\psi_0\rangle, |\psi_1\rangle$ such that $U|0\rangle = \sqrt{1-p}|\psi_0\rangle + \sqrt{p}|\psi_1\rangle$ and $\Pi U|0\rangle = \sqrt{p}|\psi_1\rangle$. If $p > 0$ then the* sequential amplitude amplification *algorithm* $\mathsf{Seq\text{-}AAmp}(U, \Pi)$ *outputs the state $|\psi_1\rangle$ with probability 1. Moreover, if we let T denote the number of applications of U, $U^\dagger$ and $I - 2\Pi$ used by the algorithm, then $\mathbb{E}[T^2] \leq O(1/p)$ and $\mathbb{E}[1/T] \leq O(\sqrt{p})$.*

The next result is a generalization of Quantum Counting that corresponds to Theorems 11 and 12 in [9].

▶ **Theorem 6** (Amplitude estimation, [9]). *Let U be a unitary quantum algorithm and let Π be a projection operator. Define the number $p \in [0,1]$ such that $p = \|\Pi U|0\rangle\|^2$. Then, for any integer $n \geq 0$, the* amplitude estimation *algorithm* $\mathsf{AEst}(U, \Pi, n)$ *outputs an amplitude estimate $\widetilde{p}$ such that, $\Pr\left[|\widetilde{p} - p| \leq \frac{2\pi\sqrt{p(1-p)}}{n} + \frac{\pi^2}{n^2}\right] \geq 8/\pi^2$. The algorithm uses n applications of U, $U^\dagger$, $I - 2\Pi$ and $O(\log^2(n))$ 2-qubit quantum gates.*

We will also use a sequential version of the amplitude estimation algorithm that does not need a time parameter n as input. This result was first obtained by [9, Theorem 15]. We describe a variant with additional properties that is based on the sequential amplitude amplification algorithm.

▶ **Theorem 7** (Sequential amplitude estimation). *There exists an algorithm, called the* sequential amplitude estimation *algorithm* $\mathsf{Seq\text{-}AEst}$*, with the following properties. Let U be a unitary quantum algorithm and let Π be a projection operator. Define the number $p \in [0,1]$ such that $p = \|\Pi U|0\rangle\|^2$. Then, the algorithm* $\mathsf{Seq\text{-}AEst}(U, \Pi)$ *outputs an amplitude estimate $\widetilde{p}$ and uses a number T of applications of U, $U^\dagger$, $I - 2\Pi$ such that,*

1. *There is a universal constant $c \in (0,1)$ such that $\Pr[|\widetilde{p} - p| \leq cp] \geq 7/8$.*
2. *There is a universal constant c' such that $\mathbb{E}[T^2] = \mathbb{E}[1/\widetilde{p}] \leq c'/p$.*
3. *There is a universal constant c'' such that $\mathbb{E}[1/T] = \mathbb{E}[\sqrt{\widetilde{p}}] \leq c''\sqrt{p}$.*

Proof. The algorithm $\mathsf{Seq\text{-}AEst}(U, \Pi)$ consists of recording the number T of applications of U, $U^\dagger$, $I - 2\Pi$ used by the sequential amplitude amplification algorithm $\mathsf{Seq\text{-}AAmp}(U, \Pi)$ (Theorem 5), and choosing the estimate $\widetilde{p} = 1/T^2$. The results follow immediately from Theorem 5 and Markov's inequality. ◀

3 Quantile estimation

In this section, we present a quantum algorithm for estimating the quantiles of a finite random variable X. This is a key ingredient for the sub-Gaussian estimator of Section 4. For the convenience of reading, we define a quantile in the following non-standard way (the cumulative distribution function is replaced with its complement).

▶ **Definition 8** (Quantile). *Given a discrete random variable X and a real $p \in [0,1]$, the* quantile *of order p is the number $Q(p) = \sup\{x \in \mathbb{R} : \Pr[X \geq x] \geq p\}$.*

Our result is inspired by the minimum finding algorithm of Dürr and Høyer [19] and its generalization in [3]. The problem of estimating the quantiles of a set of numbers under the *uniform* distribution was studied before by Nayak and Wu [34, 33]. We differ from that work by allowing arbitrary distributions, and by not using the amplitude estimation algorithm. On the other hand, we restrict ourselves to finding a constant factor estimate, whereas [34, 33] can achieve any wanted accuracy.

The idea behind our algorithm is rather simple: if we compute a sequence of values $-\infty = y_0 \leq y_1 \leq y_2 \leq y_3 \leq \dots$ where each y_{j+1} is sampled from the distribution of X conditioned on $y_{j+1} \geq y_j$, then when $j \simeq \log(1/p)$ the value of y_j should be close to the quantile $Q(p)$. The complexity of sampling each y_j is on the order of $1/\Pr[X \geq y_j]$ classically, but it can be done quadratically faster in the quantum setting. We analyze a slightly different algorithm, where the sequence of samples is strictly increasing and instead of stopping after roughly $\log(1/p)$ iterations we count the number of experiments performed by the algorithm and stop when it reaches a value close to $1/\sqrt{p}$. This requires showing that the times T_j spent on sampling y_j is neither too large nor too small with high probability, which is proved in the next lemma.

▶ **Lemma 9.** *There is a quantum algorithm such that, given a q-random variable X and a value $x \in \mathbb{R} \cup \{-\infty, +\infty\}$, it outputs a sample y from the probability distribution of X conditioned on $y > x$. If we let T denote the number of quantum experiments performed by this algorithm, then there exist two universal constants $c_0 < c_1$ such that $\mathbb{E}[T] \leq c_1/\sqrt{\Pr[X > x]}$ and $\Pr[T < c_0/\sqrt{\Pr[X > x]}] \leq 1/10$.*

Proof. Let $(\mathcal{H}, U, M)$ be a q-variable generating X. We use the comparison oracle $C_{x,+\infty}$ from Assumption 1 to construct the unitary $V = C_{x,+\infty}(U \otimes I)$ acting on $\mathcal{H} \otimes \mathbb{C}^2$. By definition of $C_{x,+\infty}$ and U (Section 2.1), we have that $V|0\rangle = \sum_{\omega \in \Omega : X(\omega) \leq x} \sqrt{p(\omega)}|\omega\rangle|0\rangle + \sum_{\omega \in \Omega : X(\omega) > x} \sqrt{p(\omega)}|\omega\rangle|1\rangle = \sqrt{1 - \Pr[X > x]}|\phi_0\rangle|0\rangle + \sqrt{\Pr[X > x]}|\phi_1\rangle|1\rangle$ for some unit states $|\phi_0\rangle, |\phi_1\rangle$ where $|\phi_1\rangle = \frac{1}{\sqrt{\Pr[X>x]}} \sum_{\omega : X(\omega) > x} \sqrt{p(\omega)}|\omega\rangle$. The algorithm for sampling y conditioned on $y > x$ consists of two steps. First, we use the sequential amplitude amplification algorithm $\mathsf{Seq\text{-}AAmp}(V, I \otimes |1\rangle\langle 1|)$ from Theorem 5 on V to obtain the state $|\phi_1\rangle$. Next, we measure $|\phi_1\rangle$ according to M. The claimed properties follow directly from Theorem 5. ◀

We use the next formula for the probability that a value x occurs in the sequence $(y_j)_j$ defined before. This lemma is adapted from [19, Lemma 1].

▶ **Lemma 10** (Lemma 47 in [3]). *Let X be a discrete random variable. Consider the increasing sequence of random variables $Y_0, Y_1, Y_2, \dots$ where Y_0 is a fixed value and Y_{j+1} for $j \geq 0$ is a sample drawn from X conditioned on $Y_{j+1} > Y_j$. Then, for any $x, y \in \mathbb{R}$,*

$$\Pr[x \in \{Y_1, Y_2, \dots\} \mid Y_0 = y] = \begin{cases} \Pr[X = x \mid X \geq x] & \text{when } x > y, \\ 0 & \text{otherwise.} \end{cases}$$

The quantile estimation algorithm is described in Algorithm 1 and the analysis is detailed in the extended version of this paper.

Algorithm 1 Quantile estimation algorithm, $\mathsf{Quantile}(X, p, \delta)$.

---

1. Repeat the following steps for $i = 1, 2, \dots, \lceil 6 \log(1/\delta) \rceil$.
   - **a.** Set $y_0 = -\infty$ and initialize a counter $C = 0$ that is incremented each time a quantum experiment is performed.
   - **b.** Set $j = 1$. Repeat the following process and interrupt it when $C = c'/\sqrt{p}$ (where c' is a constant chosen in the proof of Theorem 11): sample an element y_{j+1} from X conditioned on $y_{j+1} > y_j$ by using the algorithm of Lemma 9, set $j \leftarrow j + 1$.
   - **c.** Set $\widetilde{Q}^{(i)} = y_j$.
2. Output $\widetilde{Q} = \mathrm{median}(\widetilde{Q}^{(1)}, \dots, \widetilde{Q}^{(\lceil 6 \log(1/\delta) \rceil)})$.

▶ **Theorem 11** (Quantile estimation). *Let X be a q-random variable. Given two reals $p, \delta \in (0,1)$, the approximate quantile $\widetilde{Q}$ produced by the* quantile estimation *algorithm* $\mathsf{Quantile}(X, p, \delta)$ *(Algorithm 1) satisfies $Q(p) \leq \widetilde{Q} \leq Q(cp)$ with probability at least $1-\delta$, where $c < 1$ is a universal constant. The algorithm performs $O\Big(\frac{\log(1/\delta)}{\sqrt{p}}\Big)$ quantum experiments.*

4 Sub-Gaussian estimator

In this section, we present the main quantum algorithm for estimating the mean of a random variable with a near-quadratic speedup over the classical sub-Gaussian estimators. Our result uses the following *Bernoulli estimator*, which is a well-known adaptation of the amplitude estimation algorithm to the mean estimation problem [9, 38, 32]. The Bernoulli estimator allows us to estimate the mean of the truncated random variable $X\mathbb{1}_{a<X\leq b}$ for any a, b.

▶ **Proposition 12** (Bernoulli estimator). *There exists a quantum algorithm, called the* Bernoulli estimator*, with the following properties. Let X be a q-random variable and set as input a time parameter $n \geq 0$, two range values $0 \leq a < b$, and a real $\delta \in (0,1)$ such that $n \geq \log(1/\delta)$. Then, the Bernoulli estimator* $\mathsf{BernEst}(X, n, a, b, \delta)$ *outputs a mean estimate $\widetilde{\mu}_{a,b}$ of $\mu_{a,b} = \mathbb{E}[X\mathbb{1}_{a<X\leq b}]$ such that $|\widetilde{\mu}_{a,b} - \mu_{a,b}| \leq \frac{\sqrt{b\mu_{a,b}}\log(1/\delta)}{n} + \frac{b\log(1/\delta)^2}{n^2}$. It performs $O(n)$ quantum experiments.*

Proof. Let $(\mathcal{H}, U, M)$ be a q-variable generating X. Using the rotation oracle $R_{a,b}$ from Assumption 2, we define the unitary algorithm $V = R_{a,b}(U \otimes I)$ acting on $\mathcal{H} \otimes \mathbb{C}^2$. In order to simplify notations, let us first assume that the random variable X is only distributed in the interval (a, b). Then, we have $\mu = \mu_{a,b}$ and by definition of $R_{a,b}$ and U (Section 2.1) the operator V satisfies that $V|0\rangle = \sum_{\omega\in\Omega}\sqrt{p(\omega)}|\omega\rangle\Big(\sqrt{1-\frac{X(\omega)}{b}}|0\rangle + \sqrt{\frac{X(\omega)}{b}}|1\rangle\Big) = \sqrt{1-\frac{\mu}{b}}\Big(\sum_{\omega\in\Omega}\sqrt{\frac{p(\omega)(b-X(\omega))}{b-\mu}}|\omega\rangle\Big)|0\rangle + \sqrt{\frac{\mu}{b}}\Big(\sum_{\omega\in\Omega}\sqrt{\frac{p(\omega)X(\omega)}{\mu}}|\omega\rangle\Big)|1\rangle$. Thus, there exist some unit states $|\psi_0\rangle, |\psi_1\rangle$ such that $V|0\rangle = \sqrt{1-\frac{\mu}{b}}|\psi_0\rangle + \sqrt{\frac{\mu}{b}}|\psi_1\rangle$ and $(I \otimes |1\rangle\langle 1|)V|0\rangle = \sqrt{\frac{\mu}{b}}|\psi_1\rangle$. If X takes values outside the interval (a, b) then the same result holds with $\mu_{a,b}$ in place of μ and a different definition of $|\psi_0\rangle, |\psi_1\rangle$.

Consider the output $\widetilde{v}$ of the amplitude estimation algorithm $\mathsf{AEst}\big(V, \Pi, \big\lceil\frac{2\pi n}{\log(1/\delta)}\big\rceil\big)$ (Theorem 6) where $\Pi = I \otimes |1\rangle\langle 1|$. Then, the estimate $b\widetilde{v}$ satisfies the statement of the proposition with probability $8/\pi^2$ by Theorem 6. The Bernoulli estimator consists of running $\lceil 6\log(1/\delta)\rceil$ copies of $\mathsf{AEst}\big(V, \Pi, \big\lceil\frac{2\pi n}{\log(1/\delta)}\big\rceil\big)$ and outputting the median of the results. The success probability is at least $1-\delta$ by the Chernoff bound. ◀

The Bernoulli estimator can estimate the mean of a non-negative q-random variable X by setting $a = 0$ and $b = \max X$. However, its performance is worse than that of the classical sub-Gaussian estimators when the maximum of X is large compared to its variance. Our quantum sub-Gaussian estimator (Algorithm 2) uses the Bernoulli estimator in a more subtle way, and in combination with the quantile estimation algorithm.

▶ **Theorem 13** (Sub-Gaussian estimator). *Let X be a q-random variable with mean μ and variance σ^2. Given a time parameter n and a real $\delta \in (0,1)$ such that $n \geq \log(1/\delta)$, the* sub-Gaussian estimator $\mathsf{SubGaussEst}(X, n, \delta)$ *(Algorithm 2) outputs a mean estimate $\widetilde{\mu}$ such that, $\Pr\Big[|\widetilde{\mu} - \mu| \leq \frac{\sigma\log(1/\delta)}{n}\Big] \geq 1-\delta$. The algorithm performs $O(n\log^{3/2}(n)\log\log(n))$ quantum experiments.*

Algorithm 2 Sub-Gaussian estimator, $\mathsf{SubGaussEst}(X, n, \delta)$.

1. Set $k = \log n$ and $m = dn\sqrt{\log n}\frac{\log(9k/\delta)}{\log(1/\delta)}$, where $d > 1$ is a constant chosen in the proof of Theorem 13 (if k is not an integer, round n to the next power of two).
2. Compute the median η of $\lceil 30\log(2/\delta)\rceil$ classical samples from X and define the non-negative random variables
$$Y^+ = (X-\eta)\mathbb{1}_{X\geq\eta} \quad \text{and} \quad Y^- = -(X-\eta)\mathbb{1}_{X\leq\eta}.$$
3. Compute an estimate $\widetilde{\mu}_{Y_+}$ of $\mathbb{E}[Y_+]$ and an estimate $\widetilde{\mu}_{Y_-}$ of $\mathbb{E}[Y_-]$ by executing the following steps with $Y := Y_+$ and $Y := Y_-$ respectively:
   a. Compute an estimate $\widetilde{Q}$ of the quantile of order $p = \left(\frac{\log(1/\delta)}{6n}\right)^2$ of Y with failure probability $\delta/8$ by using the quantile estimation algorithm $\mathsf{Quantile}(Y, p, \delta/8)$.
   b. Define $a_{-1} = 0$ and $a_\ell = \frac{2^\ell}{n}\widetilde{Q}$ for $\ell \geq 0$. Compute an estimate $\widetilde{\mu}_\ell$ of $\mathbb{E}[Y\mathbb{1}_{a_{\ell-1}<Y\leq a_\ell}]$ with failure probability $\delta/(9k)$ for each $0 \leq \ell \leq k$, by using the Bernoulli estimator $\mathsf{BernEst}(Y, m, a_{\ell-1}, a_\ell, \delta/(9k))$ with m quantum experiments.
   c. Set $\widetilde{\mu}_Y = \sum_{\ell=0}^k \widetilde{\mu}_\ell$.
4. Output $\widetilde{\mu} = \eta + \widetilde{\mu}_{Y_+} - \widetilde{\mu}_{Y_-}$.

Proof. First, by standard concentration inequalities, the median η computed at step 2 satisfies $|\eta - \mu| \leq 2\sigma$ with probability at least $1 - \delta/2$. Moreover, if $|\eta - \mu| \leq 2\sigma$ then $\sqrt{\mathbb{E}[(X-\eta)^2]} = \sqrt{\mathbb{E}[(X-\mu+\mu-\eta)^2]} \leq \sqrt{\mathbb{E}[(X-\mu)^2]} + |\mu - \eta| \leq 3\sigma$, by using the triangle inequality. Below we prove that for any non-negative random variable Y the estimate $\widetilde{\mu}_Y$ of $\mu_Y = \mathbb{E}[Y]$ computed at step 3 satisfies

$$|\widetilde{\mu}_Y - \mu_Y| \leq \frac{\sqrt{\mathbb{E}[Y^2]}\log(1/\delta)}{5n} \tag{4}$$

with probability at least $1-\delta/4$. Using the fact that $X = \eta + Y_+ - Y_-$ and $(X-\eta)^2 = Y_+^2 + Y_-^2$, we can conclude that

$$|\widetilde{\mu} - \mu| \leq \frac{\left(\sqrt{\mathbb{E}[Y_+^2]} + \sqrt{\mathbb{E}[Y_-^2]}\right)\log(1/\delta)}{5n} \leq \frac{\sqrt{2\mathbb{E}[(X-\eta)^2]}\log(1/\delta)}{5n} \leq \frac{\sigma\log(1/\delta)}{n}$$

with probability at least $1 - \delta$. The algorithm performs $O(\log(1/\delta)) \leq O(n)$ classical experiments during step 2, $O(\log(1/\delta)/\sqrt{p}) \leq O(n)$ quantum experiments during step 3.a, and $O(km) \leq O(n\log^{3/2}(n)\log\log(n))$ quantum experiments during step 3.b.

We now turn to the proof of Equation (4). We make the assumption that all the subroutines used in step 3 are successful, which is the case with probability at least $(1 - \delta/8)(1-\delta/(9k))^{k+1} \geq 1 - \delta/4$. First, according to Theorem 11, we have $Q(p) \leq \widetilde{Q} \leq Q(cp)$ for some universal constant c. It implies that $cp \leq \Pr[Y \geq Q(cp)] \leq \Pr[Y \geq \widetilde{Q}] \leq \mathbb{E}[Y^2]/\widetilde{Q}^2$, where the first two inequalities are by definition of the quantile function Q, and the last inequality is a standard fact. Consequently, by our choice of p,

$$\widetilde{Q} \leq \frac{6n\sqrt{\mathbb{E}[Y^2]}}{\sqrt{c}\log(1/\delta)}. \tag{5}$$

Next, we upper bound the expectation of the part of Y that is above the largest threshold $a_k = \widetilde{Q}$ considered in step 3.b. By Cauchy–Schwarz' inequality, we have $\mathbb{E}[Y\mathbb{1}_{Y>\widetilde{Q}}] \leq$

$\sqrt{\mathbb{E}[Y^2]\Pr[Y > \widetilde{Q}]}$. Moreover, by definition of Q, $\Pr[Y > \widetilde{Q}] \leq \Pr[Y > Q(p)] \leq p$. Thus,

$$\mathbb{E}[Y \mathbb{1}_{Y>\widetilde{Q}}] \leq \frac{\sqrt{\mathbb{E}[Y^2]}\log(1/\delta)}{6n}. \tag{6}$$

The expectation of Y is decomposed into the sum $\mu_Y = \sum_{\ell=0}^{k} \mu_\ell + \mathbb{E}[Y \mathbb{1}_{Y>a_k}]$, where $\mu_\ell = \mathbb{E}[Y \mathbb{1}_{a_{\ell-1}<Y\leq a_\ell}]$ is estimated at step 3.b. We have $|\widetilde{\mu}_\ell - \mu_\ell| \leq \frac{\sqrt{a_\ell \mu_\ell}\log(1/\delta)}{dn\sqrt{\log n}} + \frac{a_\ell \log(1/\delta)^2}{d^2 n^2 \log n}$ for all $0 \leq \ell \leq k$ according to Proposition 12. Thus, by the triangle inequality,

$$\begin{aligned}
|\widetilde{\mu}_Y - \mu_Y| &\leq \sum_{\ell=0}^{k} |\widetilde{\mu}_\ell - \mu_\ell| + \mathbb{E}[Y \mathbb{1}_{Y>a_k}] \\
&\leq \sum_{\ell=0}^{k} \frac{\sqrt{a_\ell \mu_\ell}\log\left(\frac{1}{\delta}\right)}{dn\sqrt{\log n}} + \sum_{\ell=0}^{k} \frac{a_\ell \log\left(\frac{1}{\delta}\right)^2}{d^2 n^2 \log n} + \mathbb{E}[Y \mathbb{1}_{Y>a_k}] \\
&\leq \frac{\widetilde{Q}\log\left(\frac{1}{\delta}\right)}{dn^2\sqrt{\log n}} + \sum_{\ell=1}^{k} \frac{\sqrt{2\mathbb{E}[Y^2 \mathbb{1}_{a_{\ell-1}<Y\leq a_\ell}]}\log\left(\frac{1}{\delta}\right)}{dn\sqrt{\log n}} + \frac{2\widetilde{Q}\log\left(\frac{1}{\delta}\right)^2}{d^2 n^2 \log n} + \mathbb{E}[Y \mathbb{1}_{Y>a_k}] \\
&\leq \frac{\sqrt{2k}\sqrt{\sum_{\ell=1}^{k} \mathbb{E}[Y^2 \mathbb{1}_{a_{\ell-1}<Y\leq a_\ell}]}\log\left(\frac{1}{\delta}\right)}{dn\sqrt{\log n}} + \frac{3\widetilde{Q}\log\left(\frac{1}{\delta}\right)^2}{dn^2\sqrt{\log n}} + \mathbb{E}[Y \mathbb{1}_{Y>a_k}] \\
&\leq \frac{\sqrt{2k}\sqrt{\mathbb{E}[Y^2]}\log\left(\frac{1}{\delta}\right)}{dn\sqrt{\log n}} + \frac{3\widetilde{Q}\log\left(\frac{1}{\delta}\right)^2}{dn^2\sqrt{\log n}} + \mathbb{E}[Y \mathbb{1}_{Y>a_k}] \\
&\leq \frac{\sqrt{2}\sqrt{\mathbb{E}[Y^2]}\log\left(\frac{1}{\delta}\right)}{dn} + \frac{18\sqrt{\mathbb{E}[Y^2]}\log\left(\frac{1}{\delta}\right)}{\sqrt{c}dn\sqrt{\log n}} + \frac{\sqrt{\mathbb{E}[Y^2]}\log\left(\frac{1}{\delta}\right)}{6n} \\
&\leq \frac{\sqrt{\mathbb{E}[Y^2]}\log\left(\frac{1}{\delta}\right)}{5n}
\end{aligned}$$

where the third step uses $a_0\mu_0 \leq a_0^2 = (\widetilde{Q}/n)^2$ and $a_\ell \mu_\ell \leq (a_\ell/a_{\ell-1})\mathbb{E}[Y^2 \mathbb{1}_{a_{\ell-1}<Y\leq a_\ell}] \leq 2\mathbb{E}[Y^2 \mathbb{1}_{a_{\ell-1}<Y\leq a_\ell}]$ when $\ell \geq 1$, the fourth step uses the Cauchy–Schwarz inequality, the sixth step uses Equations (5) and (6), and in the last step we choose $d = 600/\sqrt{c}$. ◀

5 (ϵ, δ)-Estimators

We study the (ϵ, δ)-approximation problem under two different scenarios. First, we consider the case where we know an upper bound Δ on the coefficient of variation $|\sigma/\mu|$. As a direct consequence of Theorem 13 we obtain the following estimator that subsumes a similar result shown in [22] for non-negative random variables.

▶ **Corollary 14** (Relative estimator)**.** *There exists a quantum algorithm with the following properties. Let X be a q-random variable with mean μ and variance σ^2, and set as input a value $\Delta \geq |\sigma/\mu|$ and two reals $\epsilon, \delta \in (0, 1)$. Then, the algorithm outputs a mean estimate $\widetilde{\mu}$ such that $\Pr[|\widetilde{\mu} - \mu| > \epsilon|\mu|] \leq \delta$ and it performs $\widetilde{O}\left(\frac{\Delta}{\epsilon}\log(1/\delta)\right)$ quantum experiments.*

Proof. The algorithm runs the sub-Gaussian estimator $\mathsf{SubGaussEst}\left(X, \frac{\Delta}{\epsilon}\log(1/\delta), \delta\right)$. ◀

Next, we construct a parameter-free estimator that performs $\widetilde{O}\left(\left(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\right)\log(1/\delta)\right)$ quantum experiments in expectation for any random variable distributed in $[0, 1]$. We follow an approach similar to the classical $\mathcal{AA}$ algorithm described in [17]. We first give a sequential estimator that approximates the mean with constant relative error and that performs $O(1/\sqrt{\mu})$ quantum experiments in expectation. We use the term "sequential" in reference to sequential analysis techniques. The classical counterpart of this estimator is the Stopping Rule Algorithm in [17].

▶ **Proposition 15** (Sequential Bernoulli estimator). *There is an algorithm, called the* sequential Bernoulli estimator*, with the following properties. Let X be a q-random variable distributed in $[0,1]$ with mean μ. Then, the sequential Bernoulli estimator* Seq-BernEst(X) *outputs an estimate $\widetilde{\mu}$ and performs a number T of quantum experiments such that,*

1. *There is a universal constant $c \in (0,1)$ such that $\Pr[|\widetilde{\mu} - \mu| \leq c\mu] \geq 7/8$.*
2. *There is a universal constant c' such that $\mathbb{E}[T^2] = \mathbb{E}[1/\widetilde{\mu}] \leq c'/\mu$.*
3. *There is a universal constant c'' such that $\mathbb{E}[\sqrt{\widetilde{\mu}}] \leq c''\sqrt{\mu}$.*

Proof. The algorithm is identical to the one of Proposition 12 with $a = 0$ and $b = 1$, except that the amplitude estimation algorithm is replaced with the sequential amplitude estimation algorithm (Theorem 7). The algorithm inherits the properties proved in Theorem 7. ◀

The expected number of experiments performed by the sequential Bernoulli estimator is $\mathbb{E}[T] \leq \sqrt{\mathbb{E}[T^2]} \leq 1/\sqrt{\mu}$. The output $\widetilde{\mu}$ of the sequential Bernoulli estimator can be used in the Bernoulli estimator (Proposition 12) with parameter $n = 8\log(1/\delta)/(\epsilon\sqrt{\widetilde{\mu}})$ to solve the (ϵ,δ)-approximation problem. However, the expected number of experiments performed with this approach is $O(\log(1/\delta)/(\epsilon\sqrt{\mu}))$. We propose a better algorithm with an improved dependence on ϵ. The algorithms uses the sequential Bernoulli estimator and the sub-Gaussian estimator.

Algorithm 3 Sequential (ϵ,δ)-estimator.

1. For $i = 1, \ldots, 32\log(1/\delta)$:
   a. Compute an estimate $\widetilde{\mu}_X$ of $\mu = \mathbb{E}[X]$ by using the sequential Bernoulli estimator Seq-BernEst(X) (Proposition 15).
   b. Let Y denote the random variable $(X - X')^2/2$ where X' is independent from X and identically distributed. Compute an estimate $\widetilde{\mu}_Y$ of $\mu_Y = \mathbb{E}[Y]$ by using the sequential Bernoulli estimator Seq-BernEst(Y) (Proposition 15). Stop the computation if it performs more than $\frac{c_1}{\sqrt{\epsilon\widetilde{\mu}_X}}$ quantum experiments (where c_1 is a constant chosen in the proof of Theorem 16) and set $\widetilde{\mu}_Y = 0$.
   c. Compute a second estimate $\widetilde{\mu}_X^{(i)}$ of μ by using the sub-Gaussian estimator SubGaussEst$(X, n, 15/16)$ (Theorem 13) with $n = c_2 \max\left(\frac{\sqrt{\widetilde{\mu}_Y}}{\epsilon\widetilde{\mu}_X}, \frac{1}{\sqrt{\epsilon\widetilde{\mu}_X}}\right)$ (where c_2 is a constant chosen in the proof of Theorem 16).
2. Output $\widetilde{\mu} = \mathrm{median}\left(\widetilde{\mu}_X^{(1)}, \ldots, \widetilde{\mu}_X^{(32\log(1/\delta))}\right)$.

▶ **Theorem 16** (Sequential relative estimator). *Let X be a q-random variable distributed in $[0,1]$ with mean μ and variance σ^2. Given two reals $\epsilon,\delta \in (0,1)$ the estimate $\widetilde{\mu}$ output by the* sequential relative estimator *(Algorithm 3) satisfies $\Pr[|\widetilde{\mu} - \mu| > \epsilon\mu] \leq \delta$. The algorithm performs $\widetilde{O}\big(\big(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\big)\log(1/\delta)\big)$ quantum experiments in expectation.*

Proof. We prove that, for a fixed value of i, the estimate $\widetilde{\mu}_X^{(i)}$ computed at step 1.c satisfies $\Pr\big[|\widetilde{\mu}_X^{(i)} - \mu| \leq \epsilon\mu\big] \geq 5/8$ and the number of experiments performed during its computation is $\widetilde{O}\big(\big(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\big)\big)$ in expectation. The theorem follows by the Chernoff bound and the linearity of expectation.

Let c, c', c'' denote the constants mentioned in Proposition 15, and set $c_1 = 16c'\sqrt{(1+c)}$ and $c_2 = 4(1+c)/\sqrt{1-c}$. We assume that $|\widetilde{\mu}_X - \mu| \leq c\mu$ at step 1.a, which is the case with probability at least 7/8 by Proposition 15. The analysis of steps 1.b and 1.c is split into two cases to show that $\Pr\big[|\widetilde{\mu}_X^{(i)} - \mu| \leq \epsilon\mu\big] \geq 5/8$. First, if $\sigma \leq \sqrt{\epsilon\mu}$, then we can ignore step 1.b and consider the second term in the max at step 1.c. By Theorem 13, the estimate $\widetilde{\mu}_X^{(i)}$ satisfies $|\widetilde{\mu}_X^{(i)} - \mu| \leq \frac{4\sigma}{c_2/\sqrt{\epsilon\widetilde{\mu}_X}} \leq \frac{4\sqrt{1+c}}{c_2}\epsilon\mu \leq \epsilon\mu$ with probability 15/16. Secondly, if $\sigma \geq \sqrt{\epsilon\mu}$, then by Proposition 15 and the fact that $\mu_Y = \sigma^2$, the estimate $\widetilde{\mu}_Y$ computed at step 1.b satisfies $|\widetilde{\mu}_Y - \sigma^2| \leq c\sigma^2$ with probability 7/8 if we remove the stopping condition. Since we assumed that $\widetilde{\mu}_X \leq (1+c)\mu$, the computation is interrupted if it performs more than $\frac{c_1}{\sqrt{\epsilon\widetilde{\mu}_X}} \geq \frac{c_1}{\sqrt{(1+c)\mu_Y}} = \frac{16c'}{\sqrt{\mu_Y}}$ experiments. However, by Proposition 15 and Markov's inequality, the number of experiments performed by the sequential Bernoulli estimator at step 1.b is at most $16c'/\sqrt{\mu_Y}$ with probability at least 15/16. Consequently, we can assume that $\widetilde{\mu}_Y \geq (1-c)\sigma^2$ with success probability at least $7/8 \cdot 15/16$. In this case, by considering the first term in the max at step 1.c, the estimate $\widetilde{\mu}_X^{(i)}$ satisfies $|\widetilde{\mu}_X^{(i)} - \mu| \leq \frac{4\sigma}{c_2\sqrt{\widetilde{\mu}_Y/(\epsilon\widetilde{\mu}_X)}} \leq \frac{4(1+c)}{c_2\sqrt{1-c}}\epsilon\mu \leq \epsilon\mu$ with probability 15/16. The overall success probability is at least $(7/8)^2(15/16)^2 \geq 5/8$.

We now analyse the expected number of quantum experiments performed during the computation of $\widetilde{\mu}_X^{(i)}$. Step 1.a performs $O(1/\sqrt{\mu})$ experiments in expectation by Proposition 15. Step 1.b is stopped after $O(1/(\sqrt{\epsilon\mu}))$ experiments in expectation since $\mathbb{E}[1/\sqrt{\widetilde{\mu}_X}] \leq O(1/\sqrt{\mu})$ by Proposition 15. Step 1.c performs $\widetilde{O}\Big(\max\Big(\frac{\sqrt{\widetilde{\mu}_Y}}{\epsilon\widetilde{\mu}_X}, \frac{1}{\sqrt{\epsilon\widetilde{\mu}_X}}\Big)\Big)$ experiments by Theorem 13. The estimates $\widetilde{\mu}_Y$ and $\widetilde{\mu}_X$ are independent if we ignore the stopping condition at step 1.b, in which case $\mathbb{E}\left[\frac{\sqrt{\widetilde{\mu}_Y}}{\widetilde{\mu}_X}\right] = \mathbb{E}\left[\frac{1}{\widetilde{\mu}_X}\right]\mathbb{E}[\sqrt{\widetilde{\mu}_Y}] \leq O\left(\frac{\sigma}{\mu}\right)$ by Proposition 15. The stopping condition can only decrease this quantity. Thus, step 1.c performs $\widetilde{O}(\max(\frac{\sigma}{\epsilon\mu}, \frac{1}{\sqrt{\epsilon\mu}}))$ experiments in expectation. ◀

6 Lower bounds

We prove several lower bounds for the mean estimation problem under different scenarios. In Section 6.1, we study the number of experiments that must be performed to estimate the mean with a sub-Gaussian error rate. In Section 6.2, we study the number of experiments needed to solve the (ϵ, δ)-approximation problem. Finally, in Section 6.3, we consider the mean estimation problem in the state-based model, where the input consists of several copies of a quantum state encoding a distribution.

6.1 Sub-Gaussian estimation

We show that the quantum sub-Gaussian estimator described in Theorem 13 is optimal up to a polylogarithmic factor. We make use of the following lower bound for Quantum Search in the small-error regime.

▶ **Proposition 17** (Theorem 4 in [13]). *Let $N > 0$, $1 \leq K \leq 0.9N$ and $\delta \geq 2^{-N}$. Let $T(N, K, \delta)$ be the minimum number of quantum queries any algorithm must use to decide with failure probability at most δ whether a function $f : [N] \to \{0,1\}$ has 0 or K preimages of 1. Then, $T(N, K, \delta) \geq \Omega(\sqrt{N/K}\log(1/\delta))$.*

We construct two particular probability distributions that allow us to reduce the Quantum Search problem to the sub-Gaussian mean estimation problem.

▶ **Theorem 18.** *Let $n > 1$ and $\delta \in (0,1)$ such that $n \geq 2\log(1/\delta)$. Fix $\sigma > 0$ and consider the family $\mathcal{P}_\sigma$ of all q-random variables with variance σ^2. Let $T(n, \sigma, \delta)$ be the minimum number of quantum experiments any algorithm must perform to compute with failure probability at most δ a mean estimate $\widetilde{\mu}$ such that $|\widetilde{\mu} - \mu| \leq \frac{\sigma\log(1/\delta)}{n}$ for any $X \in \mathcal{P}_\sigma$ with mean μ. Then, $T(n, \sigma, \delta) \geq \Omega(n)$.*

Proof. Let $m = \frac{n}{\log(1/\delta)}$ and $b = \frac{m}{\sqrt{1-1/m^2}}\sigma$. We define the probability distribution p_0 with support $\{0, b\}$ that takes value b with probability $\frac{1}{m^2}$. Similarly, we define the probability distribution p_1 with support $\{0, -b\}$ that takes value $-b$ with probability $\frac{1}{m^2}$. The variance of each distribution is equal to σ^2. Moreover, the means μ_0 and μ_1 of the two distributions satisfy that,

$$\mu_0 - \mu_1 > 2\frac{\sigma \log(1/\delta)}{n}. \tag{7}$$

Let N, K be two integers such that $N \geq \log(1/\delta)$ and $K/N = 1/m^2$ (assuming m is rational). Let F_0 be the family of all functions $f : [N] \to \{0, 1\}$ with exactly K preimages of 1. Similarly, let F_1 be the family of all functions $f : [N] \to \{-1, 0\}$ with exactly K preimages of -1. By using Proposition 17, it is easy to see that any algorithm that can distinguish between $f \in F_0$ and $f \in F_1$ with success probability $1 - \delta$ must use at least $\Omega(\sqrt{N/K}\log(1/\delta)) = \Omega(m\log(1/\delta)) = \Omega(n)$ quantum queries to f. We associate with each function $f \in F_0 \cup F_1$ the q-variable $(\mathcal{H}, U, M)_f$ where $\mathcal{H} = \mathbb{C}^{N+2}$, $U|0\rangle = \frac{1}{\sqrt{N}}\sum_{x\in[N]}|x\rangle|f(x)\rangle$, and $M = \{I \otimes |0\rangle\langle 0|, I \otimes |-1\rangle\langle -1|, I \otimes |1\rangle\langle 1|\}$. The random variable X generated by $(\mathcal{H}, U, M)_f$ is distributed according to p_0 if $f \in F_0$, and according to p_1 if $f \in F_1$. Moreover, one quantum experiment with respect to X can be simulated with one quantum query to f. Consequently, any algorithm that can distinguish between a random variable distributed according to p_0 or p_1 with success probability $1-\delta$ must perform at least $\Omega(n)$ quantum experiments. On the other hand, by Equation (7), if an algorithm can estimate the mean with an error rate smaller than $\frac{\sigma\log(1/\delta)}{n}$ then it can distinguish between $f \in F_0$ and $f \in F_1$. Thus, $T(n, \sigma, \delta) \geq \Omega(n)$. ◀

6.2 (ϵ, δ)-Estimation

We consider the (ϵ, δ)-estimation problem in the parameter-free setting, when the coefficient of variation is unknown. We make use of the next lower bound for Quantum Counting.

▶ **Proposition 19** (Theorem 4.2.6 in [33]). *Let $N > 0$, $1 < K \leq N$ and $\epsilon \in \left(\frac{1}{4K}, 1\right)$. Consider the set of all quantum algorithms such that, given a query oracle to any function $f : [N] \to \{0, 1\}$, they return an estimate $\widetilde{C}$ of the number C of preimages of 1 in f such that $|\widetilde{C} - C| \leq \epsilon C$ with probability at least 2/3. Let $T_K(N, \epsilon)$ be the minimum number of quantum queries any such algorithm must use when the oracle has exactly K preimages of 1. Then, $T_K(N, \epsilon) \geq \Omega\left(\frac{\sqrt{K(N-K)}}{\epsilon K+1} + \sqrt{\frac{N}{\epsilon K+1}}\right)$.*

We obtain by a simple reduction to the above problem that the result described in Theorem 16 is nearly optimal.

▶ **Proposition 20.** *Let $\epsilon \in (0, 1)$. Let $\mathcal{P}_\mathcal{B}$ denote the family of all q-random variables that follow a Bernoulli distribution. Consider any algorithm that takes as input $X \in \mathcal{P}_\mathcal{B}$ and that outputs a mean estimate $\widetilde{\mu}$ such that $|\widetilde{\mu} - \mathbb{E}[X]| \leq \epsilon\mathbb{E}[X]$ with probability at least 2/3. Then, for any $\mu \in (0, 1)$, there exists $X \in \mathcal{P}_\mathcal{B}$ with mean μ such that the algorithm performs at least $\Omega\left(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\right)$ quantum experiments on input X, where $\sigma^2 = \mathrm{Var}[X]$.*

Proof. Given $\epsilon \in (0, 1)$ and $\mu \in (0, 1)$, we choose two integers K and N such that $K > 1/(4\epsilon)$ and $K/N = \mu$ (assuming μ is rational). Similarly to the proof of Theorem 18, we associate with each function $f : [N] \to \{0, 1\}$ the q-variable $(\mathcal{H}, U, M)_f$ where $\mathcal{H} = \mathbb{C}^{N+2}$, $U|0\rangle = \frac{1}{\sqrt{N}}\sum_{x\in[N]}|x\rangle|f(x)\rangle$, and $M = \{I \otimes |0\rangle\langle 0|, I \otimes |1\rangle\langle 1|\}$. If an algorithm can estimate the mean of any Bernoulli random variable with error ϵ and success probability $2/3$, then it can be used to count the number of preimages of 1 in f with the same accuracy. Thus, by Proposition 19, it must perform at least $\Omega\Big(\frac{\sqrt{K(N-K)}}{\epsilon K+1} + \sqrt{\frac{N}{\epsilon K+1}}\Big) = \Omega\Big(\frac{\sqrt{\mu(1-\mu)}}{\epsilon\mu+1/N} + \frac{1}{\sqrt{\epsilon\mu+1/N}}\Big) = \Omega\big(\frac{\sigma}{\epsilon\mu} + \frac{1}{\sqrt{\epsilon\mu}}\big)$ quantum experiments on a q-random variable with mean μ and variance $\sigma^2 = \mu(1-\mu)$. ◀

6.3 State-based estimation

We consider the *state-based* model where the input consists of several copies of a quantum state $|p\rangle = \sum_{x\in E}\sqrt{p(x)}|x\rangle$ encoding a distribution p over E. This model is weaker than the one described before, since it does not provide access to a unitary algorithm preparing $|p\rangle$. We prove that no quantum speedup is achievable in this setting. Our result uses the next lower bound on the number of copies needed to distinguish two states.

▶ **Lemma 21.** *Let $\delta \in (0,1)$ and consider two probability distributions p_0 and p_1 with the same finite support $E \subset \mathbb{R}$. Define the states $|\phi_0\rangle = \sum_{x\in E}\sqrt{p_0(x)}|x\rangle$ and $|\phi_1\rangle = \sum_{x\in E}\sqrt{p_1(x)}|x\rangle$. Then, the smallest integer T such that there is an algorithm that can distinguish $|\phi_0\rangle^{\otimes T}$ from $|\phi_1\rangle^{\otimes T}$ with success probability at least $1-\delta$ satisfies $T \geq \frac{\ln(1/(4\delta))}{D(p_0\|p_1)}$, where $D(p_0\|p_1) = \sum_{x\in E} p_0(x)\ln\Big(\frac{p_0(x)}{p_1(x)}\Big)$ is the KL-divergence from p_0 to p_1.*

Proof. According to Helstrom's bound [25] the best success probability to distinguish between two states $|\phi\rangle$ and $|\phi'\rangle$ is $\frac{1}{2}(1+\sqrt{1-|\langle\phi\,|\,\phi'\rangle|^2})$. Thus, the smallest number T needed to distinguish $|\phi_0\rangle^{\otimes T}$ from $|\phi_1\rangle^{\otimes T}$ must satisfy $\frac{1}{2}(1+\sqrt{1-\langle\phi_0\,|\,\phi_1\rangle^{2T}}) \geq 1-\delta$. It implies that $T \geq \frac{-\ln\left(1-(1-2\delta)^2\right)}{-2\ln(\langle\phi_0\,|\,\phi_1\rangle)} \geq \frac{\ln(1/(4\delta))}{-2\ln\left(\sum_x p_0(x)\sqrt{\frac{p_1(x)}{p_0(x)}}\right)} \geq \frac{\ln(1/(4\delta))}{\sum_x p_0(x)\ln\left(\frac{p_0(x)}{p_1(x)}\right)} = \frac{\ln(1/(4\delta))}{D(p_0\|p_1)}$ where the second inequality uses the concavity of the logarithm function. ◀

We use the above lemma to show that no quantum mean estimator can perform better than the classical sub-Gaussian estimators in the state-based input model.

▶ **Theorem 22.** *Let $n > 1$ and $\delta \in (0,1)$ such that $n \geq 2\log(1/\delta)$. Fix $\sigma > 0$ and consider the family $\mathcal{P}_\sigma$ of all distributions with finite support whose variance lies in the interval $[\sigma^2, 4\sigma^2]$. For any $p \in \mathcal{P}_\sigma$ with support $E \subset \mathbb{R}$, define the state $|p\rangle = \sum_{x\in E}\sqrt{p(x)}|x\rangle$. Let $T(n,\sigma,\delta)$ be the smallest integer such that there exists an algorithm that receives the state $|p\rangle^{\otimes T(n,\sigma,\delta)}$ for any $p \in \mathcal{P}_\sigma$, and that outputs an estimate $\widetilde{\mu}$ of the mean μ of p such that $\Pr\Big[|\widetilde{\mu}-\mu| > \sqrt{\frac{\sigma^2\log(1/\delta)}{n}}\Big] \leq \delta$. Then, $T(n,\sigma,\delta) \geq \Omega(n)$.*

Proof. Let $m = \frac{n}{\log(1/\delta)}$, $b = \frac{m}{\sqrt{m-1}}\sigma$ and $\alpha = 2\ln\Big(1+\sqrt{1-\frac{1}{m}}\Big)$. We define the two distributions p_0 and p_1 with support $E = \{0, b\}$ such that $p_0(b) = \frac{e^\alpha}{m}$ and $p_1(b) = \frac{1}{m}$. Let μ_0 and σ_0^2 (resp. μ_1 and σ_1^2) denote the expectation and the variance of p_0 (resp. p_1). Observe that $p_0, p_1 \in \mathcal{P}_\sigma$ since $\sigma_0 \in [\sigma, 2\sigma]$ and $\sigma_1 = \sigma$. Moreover, $\mu_0 - \mu_1 = \sigma\frac{e^\alpha-1}{\sqrt{m-1}} = \sigma\big(e^{\alpha/2}+1\big)\frac{e^{\alpha/2}-1}{\sqrt{m-1}} > 2\sqrt{\frac{\sigma^2\log(1/\delta)}{n}}$. Thus, we can distinguish $|p_0\rangle^{\otimes T(n,\sigma,\delta)}$ from $|p_1\rangle^{\otimes T(n,\sigma,\delta)}$ with failure probability δ by using any optimal algorithm that satisfies the error bound stated in the theorem. Since the KL-divergence from p_0 to p_1 is $D(p_0\|p_1) \leq p_0(b)\ln\Big(\frac{p_0(b)}{p_1(b)}\Big) = \frac{\alpha e^\alpha}{m^2} \leq \frac{6}{m}$, we must have $T(n,\sigma,\delta) \geq \Omega\Big(\frac{\log(1/\delta)}{D(p_1\|p_0)}\Big) = \Omega(n)$ by Lemma 21. ◀

References

1 D. S. Abrams and C. P. Williams. Fast quantum algorithms for numerical integrals and stochastic processes, 1999. `arXiv:quant-ph/9908083`.

2 N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

3 J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf. Quantum SDP-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020.

4 S. Arunachalam, A. Belovs, A. M. Childs, R. Kothari, A. Rosmanis, and R. de Wolf. Quantum coupon collector. In *Proceedings of the 15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC)*, pages 10:1–10:17, 2020.

5 S. Arunachalam and R. de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19(1):2879–2878, 2018.

6 P. J. Bickel. On some robust estimates of location. *The Annals of Mathematical Statistics*, 36(3):847–858, 1965.

7 M. Boyer, G. Brassard, P. Høyer, and A. Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4-5):493–505, 1998.

8 G. Brassard, F. Dupuis, S. Gambs, and A. Tapp. An optimal quantum algorithm to approximate the mean and its application for approximating the median of a set of points over an arbitrary distance, 2011. `arXiv:1106.4267 [quant-ph]`.

9 G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.

10 S. Bravyi, A. W. Harrow, and A. Hassidim. Quantum algorithms for testing properties of distributions. *IEEE Transactions on Information Theory*, 57(6):3971–3981, 2011.

11 N. H. Bshouty and J. C. Jackson. Learning dnf over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, 28(3):1136–1153, 1999.

12 S. Bubeck, N. Cesa-Bianchi, and G. Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711—-7717, 2013.

13 H. Buhrman, R. Cleve, R. de Wolf, and C. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of the 40th Symposium on Foundations of Computer Science (FOCS)*, pages 358–368, 1999.

14 R. Canetti, G. Even, and O. Goldreich. Lower bounds for sampling algorithms for estimating the average. *Information Processing Letters*, 53(1):17–25, 1995.

15 O. Catoni. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(4):1148–1185, 2012.

16 S. Chakraborty, E. Fischer, A. Matsliah, and R. de Wolf. New results on quantum property testing. In *Proceedings of the 30th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 145–156, 2010.

17 P. Dagum, R. Karp, M. Luby, and S. Ross. An optimal algorithm for Monte Carlo estimation. *SIAM Journal on Computing*, 29(5):1484–1496, 2000.

18 L. Devroye, M. Lerasle, G. Lugosi, and R. I. Oliveira. Sub-Gaussian mean estimators. *The Annals of Statistics*, 44(6):2695–2725, 2016.

19 C. Dürr and P. Høyer. A quantum algorithm for finding the minimum, 1996. `arXiv:quant-ph/9607014`.

20 L. Gajek, W. Niemiro, and P. Pokarowski. Optimal Monte Carlo integration with fixed relative precision. *Journal of Complexity*, 29(1):4–26, 2013.

21 L. K. Grover. A framework for fast quantum mechanical algorithms. In *Proceedings of the 30th Symposium on Theory of Computing (STOC)*, pages 53–62, 1998.

22 Y. Hamoudi and F. Magniez. Quantum Chebyshev's inequality and applications. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 69:1–69:16, 2019.

23 S. Heinrich. Quantum summation with an application to integration. *Journal of Complexity*, 18(1):1–50, 2002.

24 S. Heinrich. From Monte Carlo to quantum computation. *Mathematics and Computers in Simulation*, 62(3–6):219–230, 2003.

25 C. W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, 1969.

26 M. Huber. An optimal (ϵ, δ)-randomized approximation scheme for the mean of random variables with bounded relative variance. *Random Structures & Algorithms*, 55(2):356–370, 2019.

27 M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

28 J. C. H. Lee and P. Valiant. Optimal sub-gaussian mean estimation in $\mathbb{R}$, 2020. `arXiv:2011.08384 [math.ST]`.

29 T. Li and X. Wu. Quantum query complexity of entropy estimation. *IEEE Transactions on Information Theory*, 65(5):2899–2921, 2019.

30 G. Lugosi and S. Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

31 V. Mnih, C. Szepesvári, and J.-Y. Audibert. Empirical Bernstein stopping. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pages 672–679, 2008.

32 A. Montanaro. Quantum speedup of Monte Carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015.

33 A. Nayak. *Lower Bounds for Quantum Computation and Communication.* PhD thesis, University of California, Berkeley, 1999.

34 A. Nayak and F. Wu. The quantum query complexity of approximating the median and related statistics. In *Proceedings of the 31st Symposium on Theory of Computing (STOC)*, pages 384–393, 1999.

35 A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization.* John Wiley & Sons, 1983.

36 E. Novak. Quantum complexity of integration. *Journal of Complexity*, 17(1):2–16, 2001.

37 L. J. Schulman and V. V. Vazirani. A computationally motivated definition of parametric estimation and its applications to the Gaussian distribution. *Combinatorica*, 25(4):465–486, 2005.

38 B. M. Terhal. *Quantum Algorithms and Quantum Entanglement.* PhD thesis, University of Amsterdam, 1999.

39 J. F. Traub and H. Wozniakowski. Path integration on a quantum computer. *Quantum Information Processing*, 1(5):365–388, 2002.