Quantum Chebyshev's Inequality and Applications

Yassine Hamoudi
Université de Paris, IRIF, CNRS, F-75013 Paris, France
hamoudi@irif.fr

Frédéric Magniez
Université de Paris, IRIF, CNRS, F-75013 Paris, France
magniez@irif.fr

Abstract

In this paper we provide new quantum algorithms with polynomial speed-up for a range of problems for which no such results were known, or we improve previous algorithms. First, we consider the approximation of the frequency moments F_k of order $k \geq 3$ in the multi-pass streaming model with updates (turnstile model). We design a P-pass quantum streaming algorithm with memory M satisfying a tradeoff of $P^2M = \widetilde{\mathcal{O}}\left(n^{1-2/k}\right)$, whereas the best classical algorithm requires $PM = \Theta(n^{1-2/k})$. Then, we study the problem of estimating the number m of edges and the number t of triangles given query access to an n-vertex graph. We describe optimal quantum algorithms that perform $\widetilde{\mathcal{O}}\left(\sqrt{n}/m^{1/4}\right)$ and $\widetilde{\mathcal{O}}\left(\sqrt{n}/t^{1/6} + m^{3/4}/\sqrt{t}\right)$ queries respectively. This is a quadratic speed-up compared to the classical complexity of these problems.

For this purpose we develop a new quantum paradigm that we call Quantum Chebyshev's inequality. Namely we demonstrate that, in a certain model of quantum sampling, one can approximate with *relative error* the mean of any random variable with a number of quantum samples that is linear in the ratio of the square root of the variance to the mean. Classically the dependence is quadratic. Our algorithm subsumes a previous result of Montanaro [47]. This new paradigm is based on a refinement of the Amplitude Estimation algorithm of Brassard et al. [11] and of previous quantum algorithms for the mean estimation problem. We show that this speed-up is optimal, and we identify another common model of quantum sampling where it cannot be obtained. Finally, we develop a new technique called "variable-time amplitude estimation" that reduces the dependence of our algorithm on the sample preparation time.

2012 ACM Subject Classification Theory of computation → Quantum computation theory

Keywords and phrases Quantum algorithms, approximation algorithms, sublinear-time algorithms, Monte Carlo method, streaming algorithms, subgraph counting

Digital Object Identifier 10.4230/LIPIcs.ICALP.2019.69

Category Track A: Algorithms, Complexity and Games

Related Version A full version of the paper is available at `https://arxiv.org/abs/1807.06456`.

Funding This research was supported by the French ANR project ANR-18-CE47-0010 (QUDATA) and the QuantERA ERA-NET Cofund project QuantAlgo.

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

1 Introduction

Motivations and Background. Randomization and probabilistic methods are among the most widely used techniques in modern science, with applications ranging from mathematical economics to medicine or particle physics. One of the most successful probabilistic approaches is the Monte Carlo Simulation method for algorithm design, that relies on repeated random sampling and statistical analysis to estimate parameters and functions of interest. From Buffon's needle experiment, in the eighteenth century, to the simulations of galaxy formation

or nuclear processes, this method and its variations have become increasingly popular to tackle problems that are otherwise intractable. The Markov chain Monte Carlo method [35] led for instance to significant advances for approximating parameters whose exact computation is #P-hard [39, 37, 20, 36].

The analysis of Monte Carlo Simulation methods is often based on concentration inequalities that characterize the deviation of a random variable from some parameter. In particular, the Chebyshev inequality is a key element in the design of randomized methods that estimate some target numerical value. Indeed, this inequality guarantees that the arithmetic mean of Δ^2/ϵ^2 independent samples, from a random variable with variance σ^2 and mean μ satisfying $\Delta \geq \sigma/\mu$, is an approximation of μ under relative error ϵ with high probability. This basic result is at the heart of many computational problems, such as counting via Markov chains [35, 54], estimating graph parameters [16, 25, 28, 22], testing properties of classical [29, 8, 15, 13] or quantum [12, 7] distributions, approximating the frequency moments in the data stream model [2, 46, 4].

Various quantum algorithms have been developed to speed-up or generalize classical Monte Carlo methods (e.g. sampling the stationary distributions of Markov-chains [55, 51, 19, 53, 17], estimating the expected values of observables or partition functions [41, 56, 51, 47]). The mean estimation problem (as addressed by Chebyshev's inequality) has also been studied in the *quantum sampling model*. In this model, a distribution is represented by a unitary transformation (called a *quantum sampler*) preparing a superposition over the elements of the distribution, with the amplitudes encoding the probability mass function. A *quantum sample* is defined as one execution of a quantum sampler or its inverse. The number of quantum samples needed to estimate the mean of a distribution on a bounded space $[0, B]$, with *additive* error ϵ, was proved to be $\mathcal{O}(B/\epsilon)$ [32, 10], or $\widetilde{\mathcal{O}}(\bar{\sigma}/\epsilon)$ [47] given an upper-bound $\bar{\sigma}^2$ on the variance. On the other hand, the mean estimation problem with *relative* error ϵ can be solved with $\mathcal{O}\left(\sqrt{B}/(\epsilon\sqrt{\mu})\right)$ quantum samples [11, 56]. Interestingly, this is a quadratic improvement over $\sigma^2/(\epsilon\mu)^2$ if the sample space is $\{0, B\}$ (this case maximizes the variance). Montanaro [47] posed the problem of whether this speed-up can be generalized to other distributions. He assumed that one knows an upper bound[1] Δ on $1 + \sigma/\mu$, and gave an algorithm using[2] $\widetilde{\mathcal{O}}\left(\Delta^2/\epsilon\right)$ quantum samples (thus improving the dependence on ϵ, compared to the classical setting). This result was reformulated in [43] to show that, having bounds $L \leq \mu \leq H$, it is possible to use $\widetilde{\mathcal{O}}(\Delta/\epsilon \cdot H/L)$ quantum samples. However, it is usually the case that the only upper-bound known on μ is $H = B$. In this situation, the latter algorithm is less efficient than previous works [11, 56].

Quantum Chebyshev's Inequality. Our main contribution (Theorem 10 and Theorem 11) is to show that the mean μ of any distribution with variance σ^2 can be approximated with relative error ϵ using $\widetilde{\mathcal{O}}(\Delta \cdot \log(H/L) + \Delta/\epsilon)$ quantum samples, given an upper bound Δ on $1 + \sigma/\mu$ and two bounds L, H such that $L < \mu < H$. This is an exponential improvement in H/L compared to previous works [43]. Moreover, if $\log(H/L)$ is negligible, this is a quadratic improvement over the number of classical samples needed when using the Chebyshev inequality. A corresponding lower bound is deduced from [50] (Theorem 12). We also show (Theorem 14) that no such speed-up is possible if we only had access to *copies* of the quantum state representing the distribution.

---

[1] More precisely, Δ is an upper bound on ϕ/μ where ϕ^2 is the second moment, which satisfies $\sigma/\mu \leq \phi/\mu \leq 1 + \sigma/\mu$.

[2] We use the notation $\widetilde{\mathcal{O}}(x)$ to indicate $\mathcal{O}(x \cdot \operatorname{polylog} x)$.

Our algorithm is based on *sequential analysis*. Given a threshold $b \geq 0$, we will consider the "truncated" mean $\mu_{<b}$ defined by replacing the outcomes larger than b with 0. Using standard techniques, this mean can be encoded in the amplitude of some quantum state $\sqrt{1-\mu_{<b}/b}|\psi\rangle + \sqrt{\mu_{<b}/b}|\psi^{\perp}\rangle$ (Corollary 4). We then run the *Amplitude Estimation* algorithm of Brassard et al. [11] on this state for Δ steps (i.e. with Δ quantum samples), only to see whether the estimate of $\mu_{<b}/b$ it returns is nonzero (this is our *stopping rule*). A property of this algorithm (Corollary 4 and Remark 7) guarantees that it is zero with high probability if and only if the number of quantum samples is below the inverse $\sqrt{b/\mu_{<b}}$ of the estimated amplitude. The crucial observation (Lemma 9) is that $\sqrt{b/\mu_{<b}}$ is smaller than Δ for large values of b, and it becomes larger than Δ when $b \approx \mu\Delta^2$. Thus, by repeatedly running the amplitude estimation algorithm with Δ quantum samples, and doing $\mathcal{O}(\log(H/L))$ steps of a logarithmic search on decreasing values of b, the first non-zero value is obtained when b/Δ^2 is approximately equal to μ. The precision of the result is later improved, by using more precise "truncated" means.

This algorithm is extended to cover the common situation where one knows a non-increasing function f such that $f(\mu) \geq 1 + \sigma/\mu$, instead of having explicitly $\Delta \geq 1 + \sigma/\mu$. For this purpose, we exhibit another property (Corollary 4 and Remark 6) of the amplitude estimation algorithm, namely that it always outputs a number smaller than the estimated value (up to a constant factor) with high probability. This shall be seen as a quantum equivalent of the Markov inequality. Combined with the previous algorithm, it allows us to find a value $f(\widetilde{\mu}) \geq 1 + \sigma/\mu$, with a second logarithmic search on $\widetilde{\mu}$. This result is detailed in the full version of the paper [31].

Next, we study the quantum analogue of the following standard fact: s classical samples, each taking *average* time T_{av} to be prepared, can be obtained in total average time $s \cdot T_{av}$. The notion of "average preparation time" is adapted to the quantum setting using the framework of variable-time algorithms introduced by Ambainis [3]. This captures the situation where the superposition prepared by the quantum sampler has different parts taking different times to be computed. We develop a variable-time amplitude estimation algorithm that approximates the target value efficiently in this case. We use it in place of the standard amplitude estimation technique to obtain an algorithm whose complexity depends on the average, instead of worst-case, sample preparation time. This result is detailed in the full version of the paper [31].

Applications. We describe two applications that illustrate the use of the above results. We first study the problem of approximating the frequency moments F_k of order $k \geq 3$ in the multi-pass streaming model with updates. Classically, the best P-pass algorithms with memory M satisfy $PM = \Theta\left(n^{1-2/k}\right)$ [46, 57]. We give a quantum algorithm for which $P^2M = \widetilde{\mathcal{O}}\left(n^{1-2/k}\right)$ (Theorem 18). This problem was studied before in [48], where the author obtained quantum speed-ups for F_0, F_2 and F_∞, but no significant improvement for $k \geq 3$. Similar tradeoff results are known for Disjointness ($P^2M = \widetilde{\Theta}(n)$ in the quantum streaming model [42] vs. $PM = \Theta(n)$ classically), and Dyck(2) ($P^3M = \Omega(\sqrt{n})$ [49] vs. $PM = \widetilde{\Theta}(\sqrt{n})$ [45, 14, 34]).

Our construction starts with a classical one-pass *linear sketch* streaming algorithm [46, 4] with memory polylog n, that samples (approximately) from a distribution with mean F_k and variance $\mathcal{O}\left(n^{1-2/k}F_k^2\right)$. We implement it with a quantum sampler, that needs two passes for one quantum sample. The crucial observation is that the reverse computation of a linear sketch algorithm can be done efficiently in one pass (whereas usually that would require processing the same stream but in the reverse direction).

As a second application, we study the approximation of graph parameters using neighbor, vertex-pair and degree queries. We show that the numbers m of edges and t of triangles, in an n-vertex graph, can be estimated with $\widetilde{\Theta}\left(n^{1/2}/m^{1/4}\right)$ (Theorem 19) and $\widetilde{\Theta}\left(\sqrt{n}/t^{1/6} + m^{3/4}/\sqrt{t}\right)$ (Theorem 21) quantum queries respectively. This is a quadratic speed-up over the best classical algorithms [28, 22]. The lower bounds (Theorems 20 and 22) are obtained with a property testing to communication complexity reduction method.

The number of edges is approximated by translating a classical estimator [52] into a quantum sampler. The triangle counting algorithm is more involved. We need a classical estimator [22] approximating the number t_v of adjacent triangles to any vertex v. The average sample preparation time of this estimator being small, we obtain a quadratic speed-up for estimating t_v using our mean estimation algorithm for variable-time samplers. We then diverge from the classical triangle counting algorithm of [22], that requires to set up a data structure for sampling edges uniformly in the graph. This technique seems to be an obstacle for a quadratic speed-up. We circumvent this problem by adapting instead a bucketing approach from [21] that partitions the graph's vertices according to the value of t_v. The size of each bucket is estimated using a second quantum sampler.

2 Preliminaries

2.1 Computational Model

In this paper we consider probability distributions d on some finite sample spaces $\Omega \subset \mathbb{R}^+$. We denote by $d(x)$ the probability to sample $x \in \Omega$ in the distribution d. We also make the assumption, which is satisfied for most of applications, that Ω is equipped with an efficient encoding of its elements $x \in \Omega$. In particular, we can perform quantum computations on the Hilbert space $\mathcal{H}_\Omega$ defined by the basis $\{|x\rangle\}_{x\in\Omega}$. Moreover, given any two values $0 \leq a < b$, we assume the existence of a unitary $R_{a,b}$ that can perform the *Bernoulli sampling* (see below) in time polylogarithmic in b. In the rest of the paper we will neglect this complexity, including the required precision for implementing any of those unitary operators.

▶ **Definition 1.** *Given a finite space $\Omega \subset \mathbb{R}^+$ and two reals $0 \leq a < b$, an* (a,b)-Bernoulli sampler over Ω *is a unitary $R_{a,b}$ acting on $\mathcal{H}_\Omega \otimes \mathbb{C}^2$ and satisfying for all $x \in \Omega$:*

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

We say that Ω is Bernoulli samplable *if any (a,b)-Bernoulli sampler can be implemented in polylogarithmic time in b, when a, b have polylog-size encodings in b.*

The $R_{a,b}$ operation can be implemented with a controlled rotation, and is reminiscent of related works on mean estimation (e.g. [56, 10, 47]). In what follows, we always use $a = 0$ or $a = b/2$. Using these notions, we can now define what a *quantum sample* is.

▶ **Definition 2.** *Given a finite Bernoulli samplable space $\Omega \subset \mathbb{R}^+$ and a distribution d on Ω, a* (quantum) sampler $\mathcal{S}$ *for d is a unitary operator acting on $\mathcal{H}_g \otimes \mathcal{H}_\Omega$, for some Hilbert space $\mathcal{H}_g$, such that*

$$\mathcal{S}(|0\rangle|0\rangle) = \sum_{x\in\Omega} \sqrt{d(x)}|\psi_x\rangle|x\rangle$$

where $|\psi_x\rangle$ are arbitrary unit vectors. A quantum sample *is one execution of $\mathcal{S}$ or $\mathcal{S}^{-1}$ (including their controlled versions). The* output *of $\mathcal{S}$ is the random variable $v(\mathcal{S})$ obtained by measuring the x-register of $\mathcal{S}(|0\rangle|0\rangle)$. Its* mean *is denoted by $\mu_\mathcal{S}$, its* variance *by $\sigma^2_\mathcal{S}$, and its* second moment *by $\phi^2_\mathcal{S} = \mathbb{E}\left[v(\mathcal{S})^2\right]$.*

Given a non-negative random variable X and two numbers $0 \leq a \leq b$, we define the random variable $X_{a,b} = \mathrm{id}_{a,b}(X)$ where $\mathrm{id}_{a,b}(x) = x$ when $a \leq x < b$ and $\mathrm{id}_{a,b}(x) = 0$ otherwise. If $a = 0$, we let $X_{<b} = X_{0,b}$. Similarly, $X_{\geq b} = \mathrm{id}_{\geq b}(X)$ where $\mathrm{id}_{\geq b}(x) = x$ when $x \geq b$ and $\mathrm{id}_{\geq b}(x) = 0$ otherwise.

We motivate the use of a Bernoulli sampler $R_{a,b}$ by the following observation: for any sampler $\mathcal{S}$ and values $0 \leq a < b$, the modified sampler $\hat{\mathcal{S}} = (I_{\mathcal{H}_g} \otimes R_{a,b})(\mathcal{S} \otimes I_{\mathbb{C}^2})$ acting on $\mathcal{H}_{\hat{g}} \otimes \mathcal{H}_{\hat{\Omega}}$, where $\mathcal{H}_{\hat{g}} = \mathcal{H}_g \otimes \mathcal{H}_\Omega$ and $\hat{\Omega} = \{0,1\}$, generates the Bernoulli distribution $d(0) = 1 - p$, $d(1) = p$ of mean $p = \mathbb{E}\left[v(\hat{\mathcal{S}})\right] = b^{-1}\mathbb{E}\left[v(\mathcal{S})_{a,b}\right]$ (see the proof of Corollary 4). This central result will be used all along this paper.

Other Quantum Sampling Models. Instead of having access to the unitary $\mathcal{S}$, one could only have copies of the state $\sum_{x\in\Omega}\sqrt{d(x)}|\psi_x\rangle|x\rangle$ (as in [5] for instance). However, as we show in Theorem 14, the speed-up presented in this paper is impossible to achieve in this model. On another note, Aharonov and Ta-Shma [1] studied the *Qsampling* problem, which is the ability to prepare $\sum_{x\in\Omega}\sqrt{d(x)}|x\rangle$ given the decription of a classical circuit with output distribution d. This problem becomes straightforward if a garbage register ψ_x can be added (using standard reversible-computation techniques). Bravyi, Harrow and Hassidim [12] considered an oracle-based model, that is provably weaker than Qsampling, where a distribution $d = (d(1), \ldots, d(N))$ on $\Omega = [N]$ is represented by an oracle $O_d : [S] \to [N]$ (for some S), such that $d(x)$ equals the proportion of inputs $s \in [S]$ with $O_d(s) = x$. It is extended to the quantum query framework with a unitary $\mathcal{O}_d$ such that $\mathcal{O}_d|s\rangle|0\rangle = |s\rangle|O_d(s)\rangle$. It is not difficult to see that applying $\mathcal{O}_d$ on a uniform superposition gives $\sum_{x\in[N]}\sqrt{d(x)}\left(\frac{1}{\sqrt{d(x)S}}\sum_{s\in[S]:O_d(s)=x}|s\rangle\right)|x\rangle$, as required by Definition 2 (where $|\psi_x\rangle = \frac{1}{\sqrt{d(x)S}}\sum_{s\in[S]:O_d(s)=x}|s\rangle$). Finally, Montanaro [47] presented a model that is similar to ours, where he replaced the x-register of $\mathcal{S}(|0\rangle|0\rangle)$ with a k-qubit register (for some k) combined with a mapping $\phi : \{0,1\}^k \to \Omega$ where $x = \phi(s)$ is the sample associated to each $s \in \{0,1\}^k$.

2.2 Amplitude Estimation

The essential building block of this paper is the amplitude estimation algorithm [11], combined with ideas from [56, 10, 47], to estimate the modified mean $b^{-1}\mathbb{E}\left[v(\mathcal{S})_{a,b}\right]$ of a quantum sampler $\mathcal{S}$ to which a Bernoulli sampler $R_{a,b}$ has been applied. We will need the following result about amplitude estimation.

▶ **Theorem 3.** *There is a quantum algorithm* AmplEst*, called* Amplitude Estimation*, that takes as input a unitary operator U, an orthogonal projector Π, and an integer $t > 2$. The algorithm outputs an estimate $\widetilde{p} = \mathsf{AmplEst}\,(U, \Pi, t)$ of $p = \langle\psi|\Pi|\psi\rangle$, where $|\psi\rangle = U|0\rangle$, such that*

$$\begin{cases} |\widetilde{p} - p| \leq 2\pi\frac{\sqrt{p}}{t} + \frac{\pi^2}{t^2}, & \text{with probability } 8/\pi^2; \\ \widetilde{p} = 0, & \text{with probability } \frac{\sin^2(t\theta)}{t^2\sin^2(\theta)}. \end{cases}$$

and $0 \leq \theta \leq \pi/2$ satisfies $\sin(\theta) = \sqrt{p}$. It uses $\mathcal{O}\left(\log^2(t)\right)$ 2-qubit quantum gates (independent of U and Π) and makes $2t+1$ calls to (the controlled versions of) U and U^{-1}, and t calls to the reflection $I - 2\Pi$.

We now present an adaptation of the algorithms from [56, 10, 47] to estimate $b^{-1}\mathbb{E}\left[v(\mathcal{S})_{a,b}\right]$.

Input: a sampler $\mathcal{S}$ acting on $\mathcal{H}_g \otimes \mathcal{H}_\Omega$, two values (a, b), an integer t, a failure parameter $0 < \delta < 1$.
Output: an estimate $\widetilde{p} = \mathsf{BasicEst}(\mathcal{S}, (a, b), t, \delta)$ of $p = b^{-1}\mathbb{E}[v(\mathcal{S})_{a,b}]$

1. Let $U = (I_{\mathcal{H}_g} \otimes R_{a,b})(\mathcal{S} \otimes I_{\mathbb{C}^2})$ and $\Pi = I_{\mathcal{H}_g} \otimes I_{\mathcal{H}_\Omega} \otimes |1\rangle\langle 1|$.
2. For $i = 1, \ldots, \Theta(\log(1/\delta))$: compute $\widetilde{p}_i = \mathsf{AmplEst}(U, \Pi, t)$.
3. Output $\widetilde{p} = \mathrm{median}\{\widetilde{p}_1, \ldots, \widetilde{p}_{\Theta(\log(1/\delta))}\}$.

Algorithm 1 The Basic Estimation algorithm BasicEst.

▶ **Corollary 4.** *Consider a quantum sampler $\mathcal{S}$ and two values $0 \leq a < b$. Denote $p = b^{-1}\mathbb{E}[v(\mathcal{S})_{a,b}]$. Given an integer $t > 2$ and a real $0 < \delta < 1$, $\mathsf{BasicEst}(\mathcal{S}, (a, b), t, \delta)$ (see Algorithm 1) uses $\mathcal{O}(t \log(1/\delta))$ quantum samples and outputs $\widetilde{p}$ satisfying all of the following inequalities with probability $1 - \delta$:*

(1) $|\widetilde{p} - p| \leq 2\pi\frac{\sqrt{p}}{t} + \frac{\pi^2}{t^2}$, *for any t;* (2) $\widetilde{p} \leq (1 + 2\pi)^2 \cdot p$, *for any t;*

(3) $\widetilde{p} = 0$, *when $t < \frac{1}{2\sqrt{p}}$;* (4) $|\widetilde{p} - p| \leq \epsilon \cdot p$, *when $t \geq \frac{8}{\epsilon\sqrt{p}}$ and $0 < \epsilon < 1$.*

Proof. We show that each $\widetilde{p}_i$ satisfies the inequalities stated in the corollary, with probability $8/\pi^2$. Since $\widetilde{p}$ is the median of $\Theta(\log 1/\delta)$ such values, the probability is increased to $1 - \delta$ using the Chernoff bound.

For each $x \in \Omega$, denote $\nu_x = \frac{x}{b}$ if $a \leq x < b$, and $\nu_x = 0$ otherwise. Since $p = \sum_{x\in\Omega} \nu_x d(x)$, observe that

$$U(|0\rangle|0\rangle|0\rangle) = \sum_{x\in\Omega} \sqrt{d(x)}|\psi_x\rangle|x\rangle \left(\sqrt{1-\nu_x}|0\rangle + \sqrt{\nu_x}|1\rangle\right) = \sqrt{1-p}|\psi_0'\rangle|0\rangle + \sqrt{p}|\psi_1'\rangle|1\rangle$$

where $|\psi_0'\rangle = \frac{1}{\sqrt{1-p}} \sum_{x\in\Omega} \sqrt{d(x)}\sqrt{1-\nu_x}|\psi_x\rangle|x\rangle$ and $|\psi_1'\rangle = \frac{1}{\sqrt{p}} \sum_{x\in\Omega} \sqrt{d(x)}\sqrt{\nu_x}|\psi_x\rangle|x\rangle$ are unit vectors. Thus, the output $\widetilde{p}_i$ of the AmplEst algorithm applied on U and Π is an estimate of p satisfying the output conditions of Theorem 3. Therefore $|\widetilde{p}_i - p| \leq 2\pi\frac{\sqrt{p}}{t} + \frac{\pi^2}{t^2}$ with probability $8/\pi^2$, for any t. By plugging $t \geq \frac{8}{\epsilon\sqrt{p}}$ into this inequality we have $|\widetilde{p}_i - p| \leq \epsilon \cdot p$. By plugging $t \geq \frac{1}{2\sqrt{p}}$ we also have $|\widetilde{p}_i - p| \leq (4\pi + 4\pi^2)p$, and thus $\widetilde{p}_i \leq (1 + 2\pi)^2 \cdot p$. Finally, if $t < \frac{1}{2\sqrt{p}}$, denote $0 \leq \theta \leq \pi/2$ such that $\sin(\theta) = \sqrt{p}$ and observe that $\theta \leq \frac{\pi}{2}\sqrt{p} \leq \frac{\pi}{4t}$ (since $\frac{2}{\pi}x \leq \sin(x) \leq x$, for $x \in [0, \pi/2]$). The probability to obtain $\widetilde{p}_i = 0$ is $\frac{\sin^2(t\theta)}{t^2\sin^2(\theta)} \geq \frac{\sin^2(t\pi/(4t))}{t^2\sin^2(\pi/(4t))} \geq \frac{\sin^2(\pi/4)}{t^2(\pi/(4t))^2} = 8/\pi^2$, since $x \mapsto \sin^2(tx)/(t^2\sin^2(x))$ is decreasing for $0 < x \leq \pi/t$. Moreover, when $t < \frac{1}{2\sqrt{p}}$, the first two inequalities are obviously satisfied if $\widetilde{p}_i = 0$. ◀

The four results on p in Corollary 4 lie at the heart of this paper. We make a few comments on them.

▶ Remark 5. Consider a sampler $\mathcal{S}$ over $\Omega = \{0, 1\}$ for the Bernoulli distribution of parameter p. Using the Chebyshev inequality, we get that $\mathcal{O}\left((1-p)/(\epsilon^2 p)\right)$ classical samples are enough for estimating p with relative error ϵ. The inequality (4) of Corollary 4 shows that $t = \mathcal{O}\left(1/(\epsilon\sqrt{p})\right)$ quantum samples are sufficient. Our main result (Section 3) generalizes this quadratic speed-up to the non-Bernoulli case.

▶ Remark 6. The inequality (2) shall be seen as an equivalent of the Markov inequality[3], namely that $\widetilde{p}$ does not exceed p by a large factor with large probability.

[3] The Markov inequality for a non-negative random variable X states that $\mathbb{P}(X \geq k\mathbb{E}[X]) \leq 1/k$ for any $k > 0$. Here, although we do not need this result, it is possible to prove that $\mathbb{P}(\widetilde{p} \geq kp) \leq C/\sqrt{k}$, for some absolute constant C.

Input: a sampler $\mathcal{S}$, an integer $\Delta_{\mathcal{S}}$, two values $0 < L < H$, two reals $0 < \epsilon, \delta < 1/2$.
Output: an estimate $\widetilde{\mu}_{\mathcal{S}}$ of $\mu_{\mathcal{S}}$.

1. Set $M = 8H$ and $\widetilde{p} = 0$
2. While $\widetilde{p} = 0$ and $M \geq 2L$:
   a. Set $M = M/2$.
   b. Compute $\widetilde{p} = \mathsf{BasicEst}\left(\mathcal{S}, (0, M\Delta_{\mathcal{S}}^2), 25\Delta_{\mathcal{S}}, \delta'\right)$ where $\delta' = \frac{\delta}{2(3+\log(H/L))}$.
3. If $M < 2L$ then output $\widetilde{\mu}_{\mathcal{S}} = 0$.
4. Else, compute $\widetilde{q} = \mathsf{BasicEst}\left(\mathcal{S}, (0, \epsilon^{-1}M\Delta_{\mathcal{S}}^2), 35^2\epsilon^{-3/2}\Delta_{\mathcal{S}}, \delta/2\right)$ and output $\widetilde{\mu}_{\mathcal{S}} = (\epsilon^{-1}M\Delta_{\mathcal{S}}^2) \cdot \widetilde{q}$.

Algorithm 2 $\epsilon-$approximation of the mean of a quantum sampler $\mathcal{S}$.

▶ Remark 7. If $p \neq 0$, inequalities (3) and (4) imply that, with large probability, $t < 8/\sqrt{p}$ when $\widetilde{p} = 0$, and $t \geq 1/(2\sqrt{p})$ when $\widetilde{p} \neq 0$. This phenomenon, at $t = \Theta(1/\sqrt{p})$, is crucially used in the next section.

3 Quantum Chebyshev's Inequality

We describe our main algorithm for estimating the mean $\mu_{\mathcal{S}}$ of any quantum sampler $\mathcal{S}$, given an upper bound $\Delta_{\mathcal{S}} \geq \phi_{\mathcal{S}}/\mu_{\mathcal{S}}$ (we recall that $\phi_{\mathcal{S}}^2 = \mathbb{E}\left[v(\mathcal{S})^2\right]$ and $\sigma_{\mathcal{S}}/\mu_{\mathcal{S}} \leq \phi_{\mathcal{S}}/\mu_{\mathcal{S}} \leq 1 + \sigma_{\mathcal{S}}/\mu_{\mathcal{S}}$). The two main tools used in this section are the BasicEst algorithm of Corollary 4, and the following lemma on "truncated" means. We recall that $X_{<b}$ (resp. $X_{\geq b}$) is defined from a non-negative random variable X by substituting the outcomes greater or equal to b (resp. less than b) with 0. Note that $X = X_{<b} + X_{\geq b}$ for all $b > 0$.

▶ **Fact 8.** *For any random variable X and real numbers $0 < a \leq b$, we have $\mathbb{E}\left[X_{a,b}\right] \leq \frac{\mathbb{E}\left[X_{a,b}^2\right]}{a}$ and $\mathbb{E}\left[X_{\geq b}\right] \leq \frac{\mathbb{E}\left[X_{\geq b}^2\right]}{b}$.*

▶ **Lemma 9.** *Let X be a non-negative random variable and $\Delta \geq \sqrt{\mathbb{E}\left[X^2\right]}/\mathbb{E}\left[X\right]$. Then, for all $c_1, c_2, M > 0$ such that $c_1 \cdot \mathbb{E}\left[X\right] \leq M \leq c_2 \cdot \mathbb{E}\left[X\right]$, we have*

$$\left(1 - \frac{1}{c_1}\right)\mathbb{E}\left[X\right] \leq \mathbb{E}\left[X_{<M\Delta^2}\right] \leq \mathbb{E}\left[X\right] \quad \text{and} \quad \sqrt{c_1}\cdot\Delta \leq \sqrt{\frac{M\Delta^2}{\mathbb{E}\left[X_{<M\Delta^2}\right]}} \leq \sqrt{c_2\left(1 - \frac{1}{c_1}\right)}\cdot\Delta$$

Proof. The first inequality is a consequence of $\mathbb{E}\left[X_{<M\Delta^2}\right] = \mathbb{E}\left[X\right] - \mathbb{E}\left[X_{\geq M\Delta^2}\right]$ and $0 \leq \mathbb{E}\left[X_{\geq M\Delta^2}\right] \leq \mathbb{E}\left[X_{\geq M\Delta^2}^2\right]/(M\Delta^2) \leq \mathbb{E}\left[X^2\right]/(M\Delta^2) \leq (1/c_1)\cdot\mathbb{E}\left[X\right]$ (using Fact 8). The second inequality is a direct consequence of the left one, and of the hypothesis $c_1 \cdot \mathbb{E}\left[X\right] \leq M \leq c_2 \cdot \mathbb{E}\left[X\right]$. ◀

Our mean estimation algorithm works in two stages. We first compute a rough estimate $M \in [2\mu_{\mathcal{S}}, 2500\mu_{\mathcal{S}}]$ with $\widetilde{\mathcal{O}}\left(\Delta_{\mathcal{S}} \cdot \log(H/L)\right)$ quantum samples (where $0 < L < \mu_{\mathcal{S}} < H$ are known bounds on $\mu_{\mathcal{S}}$). Then, we improve the accuracy of the estimate to any value ϵ, at extra cost $\widetilde{\mathcal{O}}\left(\Delta_{\mathcal{S}}/\epsilon^{3/2}\right)$.

▶ **Theorem 10.** *If $\Delta_{\mathcal{S}} \geq \phi_{\mathcal{S}}/\mu_{\mathcal{S}}$ and $L < \mu_{\mathcal{S}} < H$ then the output $\widetilde{\mu}_{\mathcal{S}}$ of Algorithm 2 satisfies $|\widetilde{\mu}_{\mathcal{S}} - \mu_{\mathcal{S}}| \leq \epsilon\mu_{\mathcal{S}}$ with probability $1 - \delta$. Moreover, for any $\Delta_{\mathcal{S}}, L, H$ it satisfies $\widetilde{\mu}_{\mathcal{S}} \leq (1 + 2\pi)^2\mu_{\mathcal{S}}$ with probability $1 - \delta$. The number of quantum samples used by the algorithm is $\mathcal{O}\left(\Delta_{\mathcal{S}} \cdot \left(\log\left(\frac{H}{L}\right)\log\left(\frac{\log(H/L)}{\delta}\right) + \epsilon^{-3/2}\log\left(\frac{1}{\delta}\right)\right)\right)$.*

Proof. Assume that $\Delta_{\mathcal{S}} \geq \phi_{\mathcal{S}}/\mu_{\mathcal{S}}$ and $L < \mu_{\mathcal{S}} < H$. We denote $p = (M\Delta_{\mathcal{S}}^2)^{-1} \cdot \mathbb{E}\left[v(\mathcal{S})_{<M\Delta_{\mathcal{S}}^2}\right]$. By Lemma 9, if $M \geq 2500\mu_{\mathcal{S}}$ then $25\Delta_{\mathcal{S}} \leq \frac{1}{2\sqrt{p}}$, and if $2\mu_{\mathcal{S}} \leq M \leq 4\mu_{\mathcal{S}}$ then $25\Delta_{\mathcal{S}} > \frac{8}{\sqrt{p}}$. Therefore, by Corollary 4, with probability $1-\delta'$, the value $\widetilde{p}$ computed at Step 2.(b) is equal to 0 when $M \geq 2500\mu_{\mathcal{S}}$, and is different from 0 when $2\mu_{\mathcal{S}} \leq M \leq 4\mu_{\mathcal{S}}$. Thus, the first time Step 2.(b) of Algorithm 2 computes $\widetilde{p} \neq 0$ happens for $M \in [2\mu_{\mathcal{S}}, 2500\mu_{\mathcal{S}}]$, with probability at least $(1-\delta')^{1+\log(4H/(2\mu_{\mathcal{S}}))} > 1-\delta/2$.

Consequently, we can assume that Step 4 is executed with $M \in [2\mu_{\mathcal{S}}, 2500\mu_{\mathcal{S}}]$, and we let $M' = M/\epsilon$. According to Lemma 9 we have $(1-\epsilon/2)\mu_{\mathcal{S}} \leq \mathbb{E}\left[v(\mathcal{S})_{<M'\Delta_{\mathcal{S}}^2}\right] \leq \mu_{\mathcal{S}}$ and $35^2\epsilon^{-3/2}\Delta_{\mathcal{S}} \geq \frac{8}{(\epsilon/2)\sqrt{q}}$, where $q = (M'\Delta_{\mathcal{S}}^2)^{-1} \cdot \mathbb{E}\left[v(\mathcal{S})_{<M'\Delta_{\mathcal{S}}^2}\right]$. Thus, according to Corollary 4, the value $\widetilde{q}$ satisfies $|\widetilde{q}-q| \leq (\epsilon/2)q$ with probability $1-\delta/2$. Using the triangle inequality, it implies $|(\epsilon^{-1}M\Delta_{\mathcal{S}}^2)\cdot\widetilde{q} - \mu_{\mathcal{S}}| \leq \epsilon\mu_{\mathcal{S}}$.

If $L \geq \mu_{\mathcal{S}}$ this may only increase the probability to stop at Step 3 and output $\widetilde{\mu}_{\mathcal{S}} = 0$. If Step 4 is executed we still have $\widetilde{\mu}_{\mathcal{S}} \leq (1+2\pi)^2\mu_{\mathcal{S}}$ with probability $1-\delta$, as a consequence of Corollary 4. ◀

In the full version [31], we improve Step 4 of Algorithm 2 to obtain the following result with (nearly) optimal dependence on ϵ.

▶ **Theorem 11.** *There is an algorithm that, given a sampler $\mathcal{S}$, an integer $\Delta_{\mathcal{S}}$, two values $0 < L < H$, and two reals $0 < \epsilon, \delta < 1$, outputs an estimate $\widetilde{\mu}_{\mathcal{S}}$ of $\mu_{\mathcal{S}}$. If $\Delta_{\mathcal{S}} \geq \phi_{\mathcal{S}}/\mu_{\mathcal{S}}$ and $L < \mu_{\mathcal{S}} < H$, it satisfies $|\widetilde{\mu}_{\mathcal{S}} - \mu_{\mathcal{S}}| \leq \epsilon\mu_{\mathcal{S}}$ with probability $1-\delta$. Moreover, for any $\Delta_{\mathcal{S}}, L, H$ it satisfies $\widetilde{\mu}_{\mathcal{S}} \leq (1+2\pi)^2\mu_{\mathcal{S}}$ with probability $1-\delta$. The number of quantum samples used by the algorithm is $\mathcal{O}\left(\Delta_{\mathcal{S}} \cdot \left(\log\left(\frac{H}{L}\right)\log\left(\frac{\log(H/L)}{\delta}\right) + \epsilon^{-1}\log^{3/2}(\Delta_{\mathcal{S}})\log\left(\frac{\log\Delta_{\mathcal{S}}}{\delta}\right)\right)\right)$.*

In Section 4, we describe an $\Omega((\Delta_{\mathcal{S}}-1)/\epsilon)$ lower bound for this mean estimation problem. Before, we present three kinds of generalizations of the above algorithms.

- **Higher moments.** Given an upper-bound $\Delta_{\mathcal{S}}^2 \geq \left(\mathbb{E}\left[v(\mathcal{S})^k\right]/\mathbb{E}\left[v(\mathcal{S})\right]^k\right)^{1/(k-1)}$ on the relative moment of order $k \geq 2$, one can easily generalize Facts 8, Lemma 9 and Theorem 11 to show that $\mu_{\mathcal{S}}$ can be estimated using $\widetilde{\mathcal{O}}\left(\Delta_{\mathcal{S}} \cdot \epsilon^{-1/(2(k-1))}\log(H/L)\log(1/\delta)\right)$ quantum samples.
- **Implicit upper bound on $\boldsymbol{\phi_{\mathcal{S}}/\mu_{\mathcal{S}}}$.** If instead of an explicit value $\Delta_{\mathcal{S}} \geq \phi_{\mathcal{S}}/\mu_{\mathcal{S}}$ we are given a non-increasing function f such that $f(\mu_{\mathcal{S}}) \geq \phi_{\mathcal{S}}/\mu_{\mathcal{S}}$, we can still estimate the mean $\mu_{\mathcal{S}}$ using $\widetilde{\mathcal{O}}\left(f(\mu_{\mathcal{S}}/c)\cdot\epsilon^{-1}\log(H/L)\log(1/\delta)\right)$ quantum samples, where $c > 1$ is an absolute constant. The proof is deferred to the full version [31] (it crucially uses the Markov-like inequality "$\widetilde{\mu}_{\mathcal{S}} \leq (1+2\pi)^2\mu_{\mathcal{S}}$" of Corollary 4 and Remark 6).
- **Time complexity and variable-time samplers.** The *time complexity* (number of quantum gates) of all above algorithms is essentially equal to the number of quantum samples multiplied by the time complexity $T_{max}(\mathcal{S})$ of the considered sampler. However, $T_{max}(\mathcal{S})$ is often much larger than the more desirable *ℓ_2-average running time* $T_{\ell_2}(\mathcal{S})$ defined by Ambainis [3] in the context of variable-time algorithms, where some branches of computation may stop earlier than the others. In the full version [31], we develop a new technique called *variable-time amplitude estimation* that improves the *time* complexity of our algorithm to $\widetilde{\mathcal{O}}\left(\Delta_{\mathcal{S}}\cdot\epsilon^{-2}T_{\ell_2}(\mathcal{S})\cdot\log^4(T_{max}(\mathcal{S}))\log(H/L)\log(1/\delta)\right)$.

The last two results are combined together in the algorithm of Theorem 21 to approximate the number of triangles in any graph.

4 Optimality and Separation Results

Using a result due to Nayak and Wu [50] on approximate counting, we can show a corresponding lower bound to Theorem 11 already in the simple case of Bernoulli variables. For this purpose, we define that an algorithm $\mathcal{A}$ solves the *Mean Estimation problem for parameters* ϵ, Δ if, for any sampler $\mathcal{S}$ satisfying $\phi_{\mathcal{S}}/\mu_{\mathcal{S}} \in [\Delta, 4\Delta]$ (the constant 4 is arbitrary), it outputs a value $\widetilde{\mu}_{\mathcal{S}}$ satisfying $|\widetilde{\mu}_{\mathcal{S}} - \mu_{\mathcal{S}}| \leq \epsilon\mu_{\mathcal{S}}$ with probability 2/3.

▶ **Theorem 12.** *Any algorithm solving the Mean Estimation problem for parameters* $0 < \epsilon < 1/5$ *and* $\Delta > 1$ *on the sample space* $\Omega = \{0,1\}$ *must use* $\Omega\left((\Delta-1)/\epsilon\right)$ *quantum samples.*

Proof. Consider an algorithm $\mathcal{A}$ solving the Mean Estimation problem for parameters $0 < \epsilon < 1/5$, $\Delta > 1$ using N quantum samples. Take two integers $0 < t < n$ large enough such that $\sqrt{2}\Delta \leq \sqrt{n/t} \leq 4\Delta$ and $\epsilon t > 1$. For any oracle $\mathcal{O} : \{1, \ldots, n\} \to \{0,1\}$, define the quantum sampler $\mathcal{S}_{\mathcal{O}}(|0\rangle|0\rangle) = \frac{1}{\sqrt{n}} \sum_{i\in[n]} |i\rangle|\mathcal{O}(i)\rangle$ and let $t_{\mathcal{O}} = |\{i \in [n] : \mathcal{O}(i) = 1\}|$. Observe that $\mu_{\mathcal{S}_{\mathcal{O}}} = \phi^2_{\mathcal{S}_{\mathcal{O}}} = t_{\mathcal{O}}/n$, and one quantum sample from $\mathcal{S}_{\mathcal{O}}$ can be implemented with one quantum query to $\mathcal{O}$.

According to [50, Corollary 1.2], any algorithm that can distinguish $t_{\mathcal{O}} = t$ from $t_{\mathcal{O}} = \lceil (1+4\epsilon)t \rceil$ makes $\Omega\left(\sqrt{n/(\epsilon t)} + \sqrt{t(n-t)/(\epsilon t)}\right) = \Omega\left((\sqrt{n/t} - 1)/\epsilon\right) = \Omega\left((\Delta-1)/\epsilon\right)$ quantum queries to $\mathcal{O}$. However, given the promise that $t_{\mathcal{O}} = t$ or $t_{\mathcal{O}} = \lceil (1+4\epsilon)t \rceil$ we can use $\mathcal{A}$ with input $\mathcal{S}_{\mathcal{O}}$, ϵ, Δ to distinguish between the two cases using N samples, that is N queries to $\mathcal{O}$. Indeed, $\phi_{\mathcal{S}_{\mathcal{O}}}/\mu_{\mathcal{S}_{\mathcal{O}}} = \sqrt{n/t_{\mathcal{O}}} \in [\Delta, 4\Delta]$ for such samplers (since $\lceil (1+4\epsilon)t \rceil \leq (1+5\epsilon)t \leq 2t$). Thus, $\mathcal{A}$ must use $N = \Omega\left((\Delta-1)/\epsilon\right)$ quantum samples. ◀

One may wonder whether the quantum speed-up presented in this paper holds if we only have access to copies of a quantum state $\sum_{x\in\Omega} \sqrt{d(x)}|\psi_x\rangle|x\rangle$ (instead of access to a unitary $\mathcal{S}$ preparing it). Below we answer this question negatively. For this purpose, we define that an algorithm $\mathcal{A}$ solves the *state-based Mean Estimation problem for parameters* ϵ, Δ if, using access to some copies of an unknown state $|d\rangle = \sum_{x\in\Omega} \sqrt{d(x)}|x\rangle$ satisfying $\phi_d/\mu_d \in [\Delta, 4\Delta]$ (where $\mu_d = \sum_x d(x)x$ and $\phi_d^2 = \sum_x d(x)x^2$), it outputs a value $\widetilde{\mu}_d$ satisfying $|\widetilde{\mu}_d - \mu_d| \leq \epsilon\mu_d$ with probability 2/3.

▶ **Lemma 13.** *Consider two distributions* d, d' *represented by the states* $|d\rangle = \sum_{x\in\Omega} \sqrt{d(x)}|x\rangle$ *and* $|d'\rangle = \sum_{x\in\Omega} \sqrt{d'(x)}|x\rangle$. *The smallest integer* T *needed to be able to discriminate* $|d\rangle^{\otimes T}$ *and* $|d'\rangle^{\otimes T}$ *with success probability* 2/3 *satisfies* $T \geq \frac{\ln(9/8)}{D(d||d')}$, *where* $D(d||d')$ *is the KL-divergence from* d *to* d'.

Proof. According to Helstrom's bound [33] the best success probability to discriminate two states $|\psi\rangle$ and $|\phi\rangle$ is $\frac{1}{2}(1 + \sqrt{1 - |\langle\psi|\phi\rangle|^2})$. Consequently, T must satisfy $\frac{1}{2}(1 + \sqrt{1 - \langle d|d'\rangle^{2T}}) \geq 2/3$, which implies

$$T \geq \frac{\ln(9/8)}{-\ln(\langle d|d'\rangle^2)} = \frac{\ln(9/8)}{-2\ln\left(\sum_x d(x)\sqrt{d'(x)/d(x)}\right)} \geq \frac{\ln(9/8)}{\sum_x d(x)\ln\left(d(x)/d'(x)\right)} = \frac{\ln(9/8)}{D(d||d')}$$

where we used the concavity of the $-\ln$ function. ◀

▶ **Theorem 14.** *Any algorithm solving the state-based Mean Estimation problem for parameters* $0 < \epsilon < 1/100$ *and* $\Delta > 1$ *on the sample space* $\Omega = \{0,1\}$ *must use* $\Omega\left((\Delta^2-1)/\epsilon^2\right)$ *copies of the input state.*

Proof. Consider an algorithm $\mathcal{A}$ solving the state-based Mean Estimation problem for parameters $0 < \epsilon < 1/100$, $\Delta > 1$ using N copies of the input state. Given any $|d\rangle = \sqrt{1-p}|0\rangle + \sqrt{p}|1\rangle$ with $\phi_d/\mu_d \in [\sqrt{6}\Delta, \sqrt{8}\Delta]$ (notice that $\mu_d = \phi_d^2 = p$ and $1-p \geq 5/6 \geq 12\epsilon$), we show how to construct a state $|d'\rangle = \sqrt{1-p'}|0\rangle + \sqrt{p'}|1\rangle$ such that

$$(1)\ (1+4\epsilon)\mu_d < \mu_{d'} < (1+24\epsilon)\mu_d\ ;\ \ (2)\ \phi_{d'}/\mu_{d'} \in [\Delta, 4\Delta]\ ;\ \ (3)\ D(d||d') \leq (12\epsilon)^2/(\Delta^2-1).$$

It is clear that $\mathcal{A}$ can be used to discriminate two such states. On the other hand, according to Lemma 13, any such algorithm muse use $N = \Omega\left(1/D(d||d')\right) = \Omega\left((\Delta^2-1)/\epsilon^2\right)$ copies of the input state.

The construction of d' is adapted from [18, Section 7]. We set $p' = pe^{\alpha(1-p)}/\psi$ where $\alpha = 12\epsilon/(1-p) < 1$ and $\psi = (1-p)e^{-\alpha p} + pe^{\alpha(1-p)}$ (so that $1-p' = (1-p)e^{-\alpha p}/\psi$). We let $\dot{\psi}$ (resp. $\ddot{\psi}$) denote the first (resp. second) derivative of ψ with respect to α. A simple calculation shows that $\mu_{d'} - \mu_d = \dot{\psi}/\psi$ and $D(d||d') = \ln\psi$. Moreover, $\sigma_{d'}^2 = \mathbb{E}_{x\sim d'}\left[(x-\mu_{d'})^2\right] = \mathbb{E}_{x\sim d'}\left[(x-\mu_d)^2\right] + 2(\mu_d - \mu_{d'})\mathbb{E}_{x\sim d'}\left[x-\mu_d\right] + (\mu_d-\mu_{d'})^2 = \mathbb{E}_{x\sim d}\left[(x-p)^2 e^{\alpha(x-p)-\ln\psi}\right] - (\mu_d-\mu_{d'})^2 = \ddot{\psi}/\psi - (\dot{\psi}/\psi)^2$.

Since $\psi = \mathbb{E}_{x\sim d}\left[e^{\alpha(x-p)}\right]$, it can be deduced from the standard inequality $1+u+u^2/3 \leq e^u \leq 1+u+u^2$ (when $|u| \leq 1$) that $1 \leq 1 + \frac{p(1-p)}{3}\cdot\alpha^2 \leq \psi \leq 1 + p(1-p)\cdot\alpha^2 \leq 2$. Consequently, $\frac{2p(1-p)}{3}\cdot\alpha \leq \dot{\psi} \leq 2p(1-p)\cdot\alpha$ and $\frac{2p(1-p)}{3} \leq \ddot{\psi} \leq 2p(1-p)$. It implies that $4\epsilon p \leq \mu_{d'} - \mu_d \leq 24\epsilon p$ and $p(1-p)/3 - (24\epsilon p)^2 \leq \sigma_{d'}^2 \leq 2p(1-p)$. Thus, $(1+4\epsilon)\mu_d \leq \mu_{d'} \leq (1+24\epsilon)\mu_d \leq \sqrt{2}\mu_d$ and $\frac{1}{6}\sigma_d^2/\mu_d^2 - (24\epsilon/\sqrt{2})^2 \leq \sigma_{d'}^2/\mu_{d'}^2 \leq 2\sigma_d^2/\mu_d^2$. Since $\sigma_{d'}^2/\mu_{d'}^2 = \phi_{d'}^2/\mu_{d'}^2 - 1$ and $\phi_d/\mu_d \in [\sqrt{6}\Delta, \sqrt{8}\Delta]$, we obtain that $\Delta \leq \frac{1}{\sqrt{6}}\phi_d/\mu_d \leq \phi_{d'}/\mu_{d'} \leq \sqrt{2}\phi_d/\mu_d \leq 4\Delta$. Finally, $D(d||d') = \ln\psi \leq p(1-p)\cdot\alpha^2 = (12\epsilon)^2 p/(1-p) \leq (12\epsilon)^2/(\Delta^2-1)$. ◀

▶ Remark 15. An intermediate version of Theorem 12 can be deduced from Theorem 14, when $\mathcal{S}$ is accessed via the *reflection oracle* $\mathcal{O}_{\mathcal{S}} = I - 2\mathcal{S}(|0\rangle|0\rangle)(\langle 0|\langle 0|)\mathcal{S}^{-1}$ only (observe that this is the case for our algorithms). Indeed, according to [38, Theorem 4], for any algorithm performing q queries to a reflection oracle $\mathcal{O} = I - 2|\phi\rangle\langle\phi|$, it is possible to remove the queries to $\mathcal{O}$ by using $\sim q^2$ copies of $|\phi\rangle$ instead.

5 Applications

We describe two applications of the Quantum Chebyshev Inequality. The first one (Section 5.1) concerns the computation of the frequency moments F_k of order $k \geq 3$ in the streaming model. We design a P-pass algorithm with quantum memory M satisfying a tradeoff of $P^2M = \widetilde{\mathcal{O}}\left(n^{1-2/k}\right)$, whereas the best algorithm with classical memory requires $PM = \Theta(n^{1-2/k})$. We then study (Section 5.2) the edge and triangle counting problems in the general graph model with quantum query access. We describe nearly optimal algorithms that approximate these parameters quadratically faster than in the classical query model.

5.1 Frequency Moments in the Multi-Pass Streaming Model

In the streaming model with update (*turnstile model*), the input is a vector $x \in \mathbb{R}^n$ obtained through a stream $\vec{u} = u_1, u_2, \ldots$ of updates. Initially, $x(0) = (0, \ldots, 0)$, and each $u_j = (i, \lambda) \in [n] \times \mathbb{R}$ modifies the i-th coordinate of $x(j)$ by adding λ to it. The goal of a *streaming algorithm* $\mathcal{T}$ is to output, at the end of the stream, some function of the final vector x while minimizing the number $M \ll n$ of memory cells. In the *multi-pass* model, the same stream is repeated for a certain number P of passes, before the algorithm outputs its result.

Input: a stream $\vec{u}$, an integer $k \geq 3$, a real $\widetilde{F}_2$, an approximation parameter $0 < \epsilon < 1$.
Output: an estimate $\widetilde{F}_k$ of the frequency moment of order k of $\vec{u}$.

1. Compute $i \in [n]$ using the streaming algorithm of Theorem 16 with input $\vec{u}$, $\epsilon/4$, $\widetilde{F}_2$.
2. Compute x_i using a second pass over $\vec{u}$.
3. Output $\widetilde{F}_2 \cdot |x_i|^{k-2}$.

Estimator 3 Frequency moment F_k of a stream.

The *frequency moment of order k* is defined, for the final vector $x = (x_1, \ldots, x_n)$, as $F_k(x) = \sum_{i \in [n]} |x_i|^k$. The problem of approximating F_k when $k \geq 3$ has been addressed first with the AMS algorithm [2], that uses $\mathcal{O}\left(n^{1-1/k}\right)$ classical memory cells in the insertion-only model (where $u_j \in [n] \times \mathbb{R}^+$). A series of works in the turnstile model culminated in optimal one-pass algorithms with memory $\Theta\left(n^{1-2/k}\right)$ [44, 26], and nearly optimal P-pass algorithms with memory $\widetilde{\Theta}\left(n^{1-2/k}/P\right)$ [46, 4, 57]. In the quantum setting, Montanaro [48] obtained a small improvement in terms of the approximation parameter ϵ only.

Our algorithm relies on a classical procedure for ℓ_2 *sampling*. Given $x \in \mathbb{R}^n$, we let $D_{q,x}$ denotes the ℓ_q *distribution* that returns $i \in [n]$ with probability $\frac{|x_i|^q}{F_q(x)}$. One can observe that the (suboptimal) AMS algorithm [2] essentially samples $i \sim D_{1,x}$ and computes $F_1 \cdot |x_i|^{k-1}$. This is an unbiased estimator for $F_k(x)$ with variance $\mathcal{O}\left(n^{1-1/k} F_k(x)^2\right)$ (thus requiring to compute $\mathcal{O}\left(n^{1-1/k}\right)$ samples in one pass). Instead, we base our algorithm on the estimator $F_2(x) \cdot |x_i|^{k-2}$ where $i \sim D_{2,x}$. It reduces the variance to $\mathcal{O}\left(n^{1-2/k} F_k(x)^2\right)$ [46], but it requires a procedure for ℓ_2 sampling. To this end, we use the following algorithm from [4] to sample from an (ϵ, δ)-*approximator* to $D_{2,x}$ (meaning that each $i \in [n]$ is sampled with a probability p_i satisfying $(1-\epsilon)\frac{|x_i|^2}{F_2(x)} - \delta \leq p_i \leq (1+\epsilon)\frac{|x_i|^2}{F_2(x)} + \delta$).

▶ **Theorem 16** ([4]). *There is a randomized streaming algorithm that, given a stream $\vec{u}$ with final vector x, a real $0 < \epsilon < 1/3$ and a value $\widetilde{F}_2$ such that $|\widetilde{F}_2 - F_2(x)| \leq (1/2) \cdot F_2(x)$, outputs a value $i \in [n]$ that is distributed according to an (ϵ, n^{-2})-approximator to $D_{2,x}$. The algorithm uses $M = \mathcal{O}\left(\epsilon^{-2} \log^3 n\right)$ classical memory cells. Moreover, each element of the stream is processed in time $T_{upd} = \mathcal{O}\left(\epsilon^{-1} \log n\right)$, and the output is computed in time $T_{rec} = \mathcal{O}\left(\epsilon^{-1} n \log n\right)$ after the last element is received.*

▶ **Proposition 17** ([46, 4]). *If we let X denote the output random variable of Estimator 3, then $\mathbb{E}[X] = (1 \pm \epsilon/2) F_k$ and $\mathrm{Var}[X] \leq \mathcal{O}\left(n^{1-2/k} F_k^2\right)$, when $|\widetilde{F}_2 - F_2| \leq (\epsilon/4) \cdot F_2$.*

It is known that any deterministic computation can be made reversible, and therefore implemented by a unitary map with a limited overhead on the time and space complexities [9]. Nonetheless, implementing naively the reverse computation of a streaming algorithm would require processing the same stream but in the *reverse* direction, which may not be always possible. This motivates our specific notion of *reversible streaming algorithms*. We say that a streaming algorithm $\mathcal{T}$ with memory size M is *reversible* if there exists a streaming algorithm $\mathcal{T}^{-1}$ with memory size M such that each computational steps of $\mathcal{T}$ and $\mathcal{T}^{-1}$ are reversible, and in addition each pass of $\mathcal{T}$ can be undone by one pass of $\mathcal{T}^{-1}$ in the *same* direction. In the full version [31] we show how to make the algorithm of Theorem 16 reversible (our result is in fact more general and holds for any *linear sketch* streaming algorithm). We combine the quantum sampler that is obtained from this result with the Quantum Chebyshev Inequality (Theorem 11) to obtain the following tradeoff.

▶ **Theorem 18.** *There is a quantum streaming algorithm that, given a stream $\vec{u}$, two integers $P \geq 1$, $k \geq 3$ and an approximation parameter $0 < \epsilon < 1$, outputs an estimate $\widetilde{F}_k$ such that $|\widetilde{F}_k - F_k| \leq \epsilon F_k$ with probability $2/3$. The algorithm uses $\widetilde{\mathcal{O}}\left(n^{1-2/k}/(\epsilon P)^2\right)$ quantum memory cells, and it makes $\widetilde{\mathcal{O}}\left(P \cdot (k \log n + \epsilon^{-1})\right)$ passes over the stream $\vec{u}$.*

Proof. We first compute, in one pass, a value $\widetilde{F}_2$ such that $|\widetilde{F}_2 - F_2| \leq (\epsilon/2)F_2$ with high probability, using [2, 48] for instance. The complexity is absorbed by the final result. Then, using the *reversible* streaming algorithm associated to Estimator 3, we can design a quantum sampler $\mathcal{S}$ using memory $M = \widetilde{\mathcal{O}}\left(\epsilon^{-2} \log^3 n\right)$ such that $\mathcal{S}(|0\rangle|0\rangle) = \sum_{r \in \{0,1\}^M} |r\rangle|\psi_r\rangle|f_r\rangle$ where each $|r\rangle$ corresponds to a different random seed for the linear sketch algorithm of Theorem 16, $|f_r\rangle$ is the output of Estimator 3, and $|\psi_r\rangle$ is some garbage state obtained when making Estimator 3 reversible. According to Proposition 17, we have $\mu_{\mathcal{S}} = (1 \pm \epsilon/2)F_k$ and $\sigma_{\mathcal{S}} \leq \mathcal{O}\left(\sqrt{n^{1-2/k}}F_k\right)$. Moreover one quantum sample can be implemented with two passes over the stream.

We "concatenate" $Q = n^{1-2/k}/P^2$ such samplers and compute the mean $\bar{f} = Q^{-1} \cdot (f_{r_1} + \cdots + f_{r_Q})$ of their results to obtain

$$\bar{\mathcal{S}}(|0\rangle|0\rangle) = \sum\nolimits_{r_1,\ldots,r_Q \in \{0,1\}^M} |r_1, \ldots, r_Q\rangle|\psi_1, \ldots, \psi_Q\rangle|f_{r_1}, \ldots, f_{r_Q}\rangle|\bar{f}\rangle.$$

This sampler satisfies $\mu_{\bar{\mathcal{S}}} = \mu_{\mathcal{S}}$ and $\sigma_{\bar{\mathcal{S}}} = \sigma_{\mathcal{S}}/\sqrt{Q} \leq \mathcal{O}\left(PF_k\right)$, and it requires two passes and memory $\bar{M} = \widetilde{\mathcal{O}}\left(Q \cdot \epsilon^{-2} \log^3 n\right)$ to be implemented. Finally, we approximate F_k by applying Theorem 11 on $\bar{\mathcal{S}}$, which uses $\widetilde{\mathcal{O}}\left(P \cdot (k \log n + \epsilon^{-1})\right)$ quantum samples. ◀

5.2 Approximating Graph Parameters in the Query Model

In this section, we consider the *general graph model* [40, 27] that provides query access to a graph $G = (V, E)$ through the following operations: (1) *degree query* (given $v \in V$, returns the degree d_v of v), (2) *neighbor query* (given $v \in V$ and i, returns the i-th neighbor of v if $i \leq d_v$, and $\perp$ otherwise), and (3) *vertex-pair query* (given $u, v \in V$, indicates if $(u, v) \in E$). This is a combination of the dense graph model (pair queries) and the bounded-degree model (neighbor and degree queries). We refer the reader to [27, Chapter 10] for a more detailed discussion about it. It can be extended to the standard quantum query framework. A quantum degree query is represented as a unitary $\mathcal{O}_{deg}$ such that $\mathcal{O}_{deg}|v\rangle|b\rangle = |v\rangle|y \oplus d_v\rangle$ where $v \in V$ and $y \in \{0,1\}^{\lceil \log n \rceil}$. The quantum neighbor $\mathcal{O}_{neigh}$ and vertex-pair $\mathcal{O}_{pair}$ queries are defined similarly. The *query complexity* of an algorithm in the *quantum general graph model* is the number of times it uses $\mathcal{O}_{deg}$, $\mathcal{O}_{nei}$ or $\mathcal{O}_{pair}$.

In the following, we let n denote the number of vertices, m the number of edges and t the number of triangles in G. We consider the problems of estimating m and t, for which we provide nearly optimal quantum algorithms. The description and analysis of these algorithms is deferred to the full version [31].

Edge counting. In the classical setting, Feige [25] showed that $\Theta(n/(\epsilon\sqrt{m}))$ degree queries are sufficient to compute a factor $(2+\epsilon)$ approximation of m, but no factor $(2-\epsilon)$ approximation can be obtained in sublinear time. Using both degree and neighbor queries, it is possible to compute a factor $(1+\epsilon)$ approximation with $\Theta\left(n/(\sqrt{\epsilon m})\right)$ classical queries [28, 52, 23]. These results were generalized to k-star counting in [30, 23]. In the quantum setting, we prove the following results.

▶ **Theorem 19.** *There is an algorithm that, given query access to any n-vertex graph G with m edges, and an approximation parameter $\epsilon < 1$, outputs an estimate $\widetilde{m}$ of m such that $|\widetilde{m} - m| \leq \epsilon m$ with probability $2/3$. This algorithm performs $\widetilde{\mathcal{O}}\left(\frac{n^{1/2}}{\epsilon m^{1/4}}\right)$ quantum degree and neighbor queries in expectation. Moreover, it does not use vertex-pair queries.*

▶ **Theorem 20.** *Any algorithm that computes an ϵ-approximation of the number m of edges in any n-vertex graph, given query access to it, must use $\Omega\left(\frac{n^{1/2}}{(\epsilon m)^{1/4}} \cdot \log^{-1}(n)\right)$ quantum queries in expectation.*

Triangle counting. In the classical general graph model, the triangle counting problem requires $\widetilde{\Theta}(n/t^{1/3} + \min(m, m^{3/2}/t))$ queries in expectation [21, 22]. This result was generalized to k-clique counting in [24]. In the quantum setting, we prove the following results.

▶ **Theorem 21.** *There is an algorithm that, given query access to any n-vertex graph G with m edges and t triangles, and an approximation parameter $\epsilon < 1$, outputs an estimate $\widetilde{t}$ of t such that $|\widetilde{t} - t| \leq \epsilon t$ with probability $2/3$. This algorithm performs $\widetilde{\mathcal{O}}\left(\left(\frac{\sqrt{n}}{t^{1/6}} + \frac{m^{3/4}}{\sqrt{t}}\right) \cdot poly(1/\epsilon)\right)$ quantum queries in expectation.*

▶ **Theorem 22.** *Any algorithm that computes an ϵ-approximation to the number t of triangles in any n-vertex graph with m vertices, given query access to it, must use $\Omega\left(\left(\frac{\sqrt{n}}{t^{1/6}} + \frac{m^{3/4}}{\sqrt{t}}\right) \cdot \log^{-1}(n)\right)$ quantum queries in expectation.*

6 Open Questions

Is it possible to improve the complexity of our main result (Theorem 11) to $\mathcal{O}(\Delta_{\mathcal{S}}/\epsilon)$ exactly? Can we generalize it to sample spaces with negative values? What are other possible applications? Two promising problems are minimum spanning tree weight [16] and arbitrary subgraph counting [24, 6].

References

1. D. Aharonov and A. Ta-Shma. Adiabatic Quantum State Generation. *SIAM Journal on Computing*, 37(1):47–82, 2007.
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. A. Ambainis. Variable time amplitude amplification and a faster quantum algorithm for solving systems of linear equations. Technical report, arXiv.org, 2010. `arXiv:1010.4458`.
4. A. Andoni, R. Krauthgamer, and K. Onak. Streaming Algorithms via Precision Sampling. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, FOCS '11, pages 363–372, 2011.
5. S. Arunachalam and R. de Wolf. Optimal Quantum Sample Complexity of Learning Algorithms. In *Proceedings of the 32nd Computational Complexity Conference*, CCC '17, pages 25:1–25:31, 2017.
6. S. Assadi, M. Kapralov, and S. Khanna. A Simple Sublinear-Time Algorithm for Counting Arbitrary Subgraphs via Edge Sampling. In *Proceedings of the 10th Conference on Innovations in Theoretical Computer Science*, ITCS '19, pages 6:1–6:20, 2019.
7. C. Badescu, R. O'Donnell, and J. Wright. Quantum state certification. Technical report, arXiv.org, 2017. `arXiv:1708.06002`.
8. T. Batu, L. Fortnow, R. Rubinfeld, W. D. Smith, and P. White. Testing Closeness of Discrete Distributions. *Journal of the ACM*, 60(1):4:1–4:25, 2013.

9 C. Bennett. Time/Space Trade-Offs for Reversible Computation. *SIAM Journal on Computing*, 18(4):766–776, 1989.

10 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. Technical report, arXiv.org, 2011. `arXiv:1106.4267`.

11 G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. *Quantum Computation and Quantum Information: A Millennium Volume*, 1:53–74, 2002.

12 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.

13 C. L. Canonne, I. Diakonikolas, D. M. Kane, and A. Stewart. Testing conditional independence of discrete distributions. In *Proceedings of the 50th Symposium on Theory of Computing*, STOC '18, pages 735–748, 2018.

14 A. Chakrabarti, G. Cormode, R. Kondapally, and A. McGregor. Information Cost Tradeoffs for Augmented Index and Streaming Language Recognition. *SIAM Journal on Computing*, 42(1):61–83, 2013.

15 S. Chan, I. Diakonikolas, P. Valiant, and G. Valiant. Optimal Algorithms for Testing Closeness of Discrete Distributions. In *Proceedings of the 25th Symposium on Discrete Algorithms*, SODA '14, pages 1193–1203, 2014.

16 B. Chazelle, R. Rubinfeld, and L. Trevisan. Approximating the Minimum Spanning Tree Weight in Sublinear Time. *SIAM Journal on Computing*, 34(6):1370–1379, 2005.

17 A. N. Chowdhury and R. D. Somma. Quantum Algorithms for Gibbs Sampling and Hitting-time Estimation. *Quantum Information and Computation*, 17(1-2):41–64, 2017.

18 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.

19 N. Destainville, B. Georgeot, and O. Giraud. Quantum Algorithm for Exact Monte Carlo Sampling. *Physical Review Letters*, 104:250502, 2010.

20 M. Dyer, A. Frieze, and R. Kannan. A Random Polynomial-time Algorithm for Approximating the Volume of Convex Bodies. *Journal of the ACM*, 38(1):1–17, 1991.

21 T. Eden, A. Levi, and D. Ron. Approximately Counting Triangles in Sublinear Time. Technical Report TR15-046, ECCC, 2015.

22 T. Eden, A. Levi, D. Ron, and C. Seshadhri. Approximately Counting Triangles in Sublinear Time. *SIAM Journal on Computing*, 46(5):1603–1646, 2017.

23 T. Eden, D. Ron, and C. Seshadhri. Sublinear Time Estimation of Degree Distribution Moments: The Degeneracy Connection. In *Proceedings of the 44th International Colloquium on Automata, Languages, and Programming*, ICALP '17, pages 7:1–7:13, 2017.

24 T. Eden, D. Ron, and C. Seshadhri. On Approximating the Number of K-cliques in Sublinear Time. In *Proceedings of the 50th Symposium on Theory of Computing*, STOC '18, pages 722–734, 2018.

25 U. Feige. On Sums of Independent Random Variables with Unbounded Variance and Estimating the Average Degree in a Graph. *SIAM Journal on Computing*, 35(4):964–984, 2006.

26 S. Ganguly. Taylor Polynomial Estimator for Estimating Frequency Moments. In *Proceedings of the 42nd International Colloquium on Automata, Languages and Programming*, ICALP '15, pages 542–553, 2015.

27 O. Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

28 O. Goldreich and D. Ron. Approximating Average Parameters of Graphs. *Random Structures & Algorithms*, 32(4):473–493, 2008.

29 O. Goldreich and D. Ron. On Testing Expansion in Bounded-Degree Graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 68–75. Springer-Verlag, 2011.

30 M. Gonen, D. Ron, and Y. Shavitt. Counting Stars and Other Small Subgraphs in Sublinear-Time. *SIAM Journal on Discrete Mathematics*, 25(3):1365–1411, 2011.

31 Y. Hamoudi and F. Magniez. Quantum Chebyshev's Inequality and Applications. Technical report, arXiv.org, 2019. `arXiv:1807.06456`.

32 S. Heinrich. Quantum Summation with an Application to Integration. *Journal of Complexity*, 18(1):1–50, 2002.

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

34 R. Jain and A. Nayak. The Space Complexity of Recognizing Well-Parenthesized Expressions in the Streaming Model: the Index Function Revisited. *IEEE Transactions on Information Theory*, 60(10):6646–6668, 2014.

35 M. Jerrum and A. Sinclair. The Markov Chain Monte Carlo Method: An Approach to Approximate Counting and Integration. In *Approximation Algorithms for NP-hard Problems*, chapter 12, pages 482–520. PWS Publishing, 1996.

36 M. Jerrum, A. Sinclair, and E. Vigoda. A Polynomial-time Approximation Algorithm for the Permanent of a Matrix with Nonnegative Entries. *Journal of the ACM*, 51(4):671–697, 2004.

37 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.

38 Z. Ji, Y.-K. Liu, and F. Song. Pseudorandom Quantum States. In *Advances in Cryptology*, CRYPTO '18, pages 126–152, 2018.

39 R. M. Karp and M. Luby. Monte-Carlo algorithms for enumeration and reliability problems. In *Proceedings of the 24th Symposium on Foundations of Computer Science*, FOCS '83, pages 56–64, 1983.

40 T. Kaufman, M. Krivelevich, and D. Ron. Tight Bounds for Testing Bipartiteness in General Graphs. *SIAM Journal on Computing*, 33(6):1441–1483, 2004.

41 E. Knill, G. Ortiz, and R. D. Somma. Optimal quantum measurements of expectation values of observables. *Physical Review A*, 75:012328, 2007.

42 F. Le Gall. Exponential Separation of Quantum and Classical Online Space Complexity. *Theory of Computing Systems*, 45(2):188–202, 2009.

43 T. Li and X. Wu. Quantum query complexity of entropy estimation. Technical report, arXiv.org, 2017. `arXiv:1710.06025`.

44 Y. Li and D. P. Woodruff. A Tight Lower Bound for High Frequency Moment Estimation with Small Error. In *Proceedings of the Workshop on Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques*, APPROX/RANDOM '13, pages 623–638, 2013.

45 F. Magniez, C. Mathieu, and A. Nayak. Recognizing Well-Parenthesized Expressions in the Streaming Model. *SIAM Journal on Computing*, 43(6):1880–1905, 2014.

46 M. Monemizadeh and D. P. Woodruff. 1-pass Relative-error Lp-sampling with Applications. In *Proceedings of the 21st Symposium on Discrete Algorithms*, SODA '10, pages 1143–1160, 2010.

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

48 A. Montanaro. The quantum complexity of approximating the frequency moments. *Quantum Information and Computation*, 16:1169–1190, 2016.

49 A. Nayak and D. Touchette. Augmented Index and Quantum Streaming Algorithms for DYCK(2). In *Proceedings of the 32nd Conference on Computational Complexity*, CCC '17, pages 23:1–23:21, 2017.

50 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 '99, pages 384–393, 1999.

51 D. Poulin and P. Wocjan. Sampling from the Thermal Quantum Gibbs State and Evaluating Partition Functions with a Quantum Computer. *Physical Review Letters*, 103:220502, 2009.

52 C. Seshadhri. A simpler sublinear algorithm for approximating the triangle count. Technical report, arXiv.org, 2015. `arXiv:1505.01927`.

53 K. Temme, T. J. Osborne, K. Vollbrecht, D. Poulin, and F. Verstraete. Quantum Metropolis Sampling. *Nature*, 471:87, 2011.

54 D. Štefankovič, S. Vempala, and E. Vigoda. Adaptive Simulated Annealing: A Near-optimal Connection Between Sampling and Counting. *Journal of the ACM*, 56(3):18:1–18:36, 2009.

55 P. Wocjan and A. Abeyesinghe. Speedup via quantum sampling. *Physical Review A*, 78:042336, 2008.

56 P. Wocjan, C.-F. Chiang, D. Nagaj, and A. Abeyesinghe. Quantum algorithm for approximating partition functions. *Physical Review A*, 80:022340, 2009.

57 D. P. Woodruff and Q. Zhang. Tight Bounds for Distributed Functional Monitoring. In *Proceedings of the 44th Symposium on Theory of Computing*, STOC '12, pages 941–960, 2012.