



A parameterized quantum circuit for estimating distribution measures

Or Peretz¹ · Michal Koren¹

Received: 21 November 2023 / Accepted: 26 March 2024
© The Author(s) 2024

Abstract

Quantum computing is a new and exciting field with the potential to solve some of the world's most challenging problems. Currently, with the rise of quantum computers, the main challenge is the creation of quantum algorithms (under the limitations of quantum physics) and making them accessible to scientists who are not physicists. This study presents a parametrized quantum circuit and its implementation in estimating the distribution measures for discrete value vectors. Various applications can be derived from this method, including information analysis, exploratory data analysis, and machine learning algorithms. This method is unique in providing access to quantum computation and enabling users to run it without prior knowledge of quantum physics. The proposed method was implemented and tested over a dataset and five discrete value distributions with different parameters. The results showed a high level of agreement between the classical computation and the proposed method using quantum computing. The maximum error obtained for the dataset was 5.996%, while for the discrete distributions, a maximum error of 5% was obtained.

Keywords Quantum computing · Parametrized quantum circuit · Data analysis · Distribution measures

1 Introduction

The field of quantum computing (QC) is considered one of the most promising in computational science (Ying 2010) and has gained importance in international research. Based on the physics theorem, QC assumes that an electron can behave simultaneously as a wave and particle (Robertson 1943). Due to the sensitivity of quantum computers to noise and decoherence (Bennett et al. 1997), it can be challenging to build and maintain a superposition of quantum computers. There have been many investigations and discussions regarding quantum computers (Zeng et al. 2017), including arguments concerning the advantages and disadvantages of quantum computers (Boyer et al. 1998). Quantum computers have the potential to revolutionize the field of computing, as they can perform tasks much faster than classical computers and can also process large amounts of data in a short amount

of time. For example, quantum computers can perform calculations that are impossible with classical computers, such as solving certain types of algorithms. Consequently, they have the potential to revolutionize many industries, such as finance, healthcare, and artificial intelligence (Piattini et al. 2021).

Quantum computers significantly reduce the complexity of computing. Due to their parallel processing capabilities, they can perform a greater variety of operations than classical computers (Biamonte et al. 2017; Wiebe 2020). Currently, there is no distinction between classical and quantum computers, and algorithms can be implemented on both (Buffoni and Caruso 2021). Hence, quantum machine learning (QML) is a young but rapidly growing field alongside QC. By using quantum gates, it is possible to transform classical machine learning algorithms into QC (Benedetti et al. 2019; Alchieri et al. 2021). Ultimately, combining classical and QC provides a powerful tool for solving complex problems.

A parameterized quantum circuit (PQC) consists of one or more parameters that can be changed according to the user's requirements (Hubregtsen et al. 2021). Their versatility makes them useful for implementing machine learning algorithms, variational quantum algorithms, and quantum simulators (Benedetti et al. 2019; Du et al. 2020). A

✉ Michal Koren
michal.koren@shenkar.ac.il

Or Peretz
or.perets@shenkar.ac.il

¹ School of Industrial Engineering and Management, Shenkar—Engineering, Design, Art, Anne Frank 12, Ramat-Gan 5252626, Israel

parameterized gate differs from a fixed gate in that it depends on variables for its operation. PQCs are a rapidly developing field, and new advances are being made constantly (Peham et al. 2023). Recent studies have described the implementation of quantum PQC and algorithms for the learning of random variables (González et al. 2022; Pirhooshyaran and Terlaky 2021) and entropy estimation (Koren et al. 2023), as well as building a quantum convolutional network to learn images (Hur et al. 2022; Tüysüz et al. 2021), implementing reinforcement learning (Dalla Pozza et al. 2022), and developing generative adversarial networks (GANs) and transfer learning (Assouel et al. 2022; Azevedo et al. 2022).

For machine learning models to be successful, data representation is crucial. Classical machine learning relies on numerical representations of data in order to be best processed by a classical algorithm. Quantum machine learning poses the same fundamental question: how to represent and efficiently input data into quantum systems to be analyzed by quantum algorithms (Li et al. 2022; Weigold et al. 2021a, b). As a result of this process, quantum machine learning algorithms are directly affected by their computational capacity (Dilip et al. 2022; LaRose and Coyle 2020; Sierra-Sosa et al. 2023). There are three primary encoding methods: (1) *Basic encoding*, which associates a classical string with a computational basis state. It is the simplest method to understand, although the state vectors quickly become sparse (i.e., vectors that have mostly zero values). (2) *Amplitude encoding* encodes data into the amplitudes of a quantum state. As a system of n qubits provides 2^n amplitudes, it can encode a dataset of N records over M features using $\log_2(N \cdot M) + 1$ qubits. (3) *Angle encoding* encodes M features into the rotation angles of M qubits. The angle encoding slightly differs from other encoding techniques as it only encodes one data point at a time, rather than an entire dataset. This method requires, at most, M qubits. All of the mentioned methods have been found to produce enhanced results in quantum autoencoders and image processing (Bravo-Prieto 2021; Majji et al. 2023; Romero et al. 2017; Shin et al. 2023).

Exploratory data analysis (EDA) is a set of techniques developed in 1970 that aims to examine the data before building a model (Tukey 1977). It explores data for patterns, trends, underlying structures, anomalies, and more (Chatfield 1986; Leinhardt and Wasserman 1979). The main goal of EDA is to develop valid models based on data insights (Komorowski et al. 2016; Morgenthaler 2009). This process can be categorized into three main categories: (1) *Univariate analysis* refers to one dependable variable (in a dataset, it explores each variable separately). It can be performed by statistical analyses, such as mean, median, and variance (Behrens and Yu 2003; Vigni et al. 2013). (2) *Bivariate analysis* examines the relationship between the two variables. It can be two numerical variables, two categorical variables, or mixed variables (Cleff 2014; Jebb et al. 2017).

(3) *Multivariate analysis* refers to at least three variables (Gelman 2004; Wang et al. 2023; Wongsuphasawat et al. 2019). Notably, analyzing the distribution of a dataset significantly impacts the results of the analysis. Furthermore, understanding the distribution can assist in choosing the appropriate statistical test, identifying anomalies, examining normality, and more, leading to results that are more accurate, reliable, and valid.

This study presents and describes a quantum-parametrized gate and its circuit implementation to estimate the distribution measures for discrete value vectors. It can be applied in information analysis, exploratory data analysis, and machine learning algorithms. There are two central innovative aspects of this proposed method: (1) a new quantum method for estimating distribution measures, such as expectation and variance, and (2) the accessibility of QC and the creation of a method that can be run without any prior knowledge of implementing quantum circuits. Section 2 will present the definition of the new quantum gate, the procedure, its implementation, and the mathematical justifications. Section 3 will describe the empirical study of a dataset and a detailed scenario of variance estimation. The method was tested and compared over five discrete value distributions, the results of which are presented in Section 4. Lastly, Section 5 will discuss the main conclusions and suggestions for future directions.

2 Quantum distribution measures

This section presents and describes a new parametrized quantum gate for statistical measure estimations. First, the general parametrized gate will be described, and its unitarity will be proven. Then, the quantum method will be presented, including the relevant logic and gates. Last, its implementation and correctness will be presented in detail.

2.1 Parameterized quantum gate

This study defines a new parameterized quantum gate that inputs $a, r \in \mathbb{N}$, creating a parameterized and diagonal square matrix of size a as follows:

$$M(a, r) = \begin{bmatrix} \sqrt{1^{ri}} & \dots & 0 & 0 \\ 0 & \sqrt{2^{ri}} & 0 & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & \dots & 0 & \sqrt{3^{ri}} \end{bmatrix}$$

A new operation defined by a quantum circuit is required to be unitary¹ since any physical operation on a state is used

¹ A matrix U is unitary if, and only if, its conjugate transpose is equal to its inverse, i.e., $U^*U = UU^* = UU^{-1} = I$.

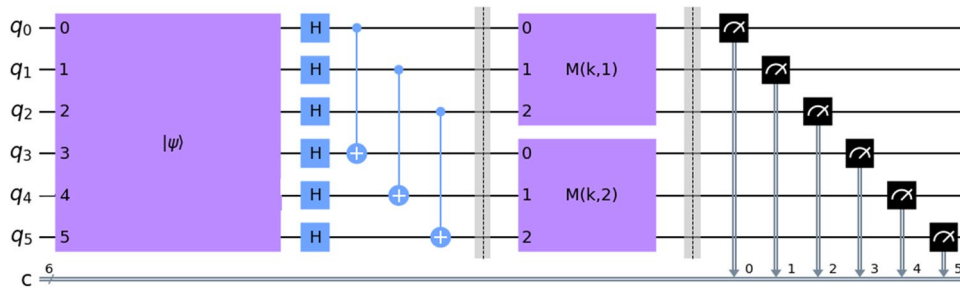


Fig. 1 The quantum circuit of variance calculation. Notes: (1) The dashed lines describe the sampling of the state in superposition. (2) IBM simulators were used (with the Qiskit Library for Python; Cross 2018) to avoid noise and sample the state vector in each time phase in the circuit. (3) Figure 1 describes the quantum circuit over several qubits, although generalization to a higher dimension can be done with tensor products. (4) The output represents an approximation of

the variance value. The analysis of the approximation ratio is detailed in Section 2.3. (5) Amplitude encoding was chosen due to the ability of this method to encode an entire dataset using a logarithmic number of qubits. Despite the number of gates required for this coding, this method can form the base for many distribution measures (e.g., expectation, variance, skewness, kurtosis). Future studies are necessary to explore a more efficient representation of these gates

to advance it (Bennett et al. 1997). Since $M(a, r)$ is a square and diagonal matrix, its inverse is also diagonal and defined as follows:

$$M(a, r)^{-1} = \begin{bmatrix} \sqrt{1^{-ri}} & \dots & 0 & 0 \\ 0 & \sqrt{2^{-ri}} & 0 & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & \dots & 0 & \sqrt{3^{-ri}} \end{bmatrix}$$

Thus, the inverse matrix is also its conjugate transpose, and the new quantum operation is unitary since:

$$M(a, r)^* M(a, r) = M(a, r)^{-1} M(a, r) = I$$

2.2 Quantum logic and gates

Let $A = (a_1, a_2, \dots, a_n)$ be the input vector, such that $\forall a_i \in A$. Therefore, it holds that $a_i \in \mathbb{N} \cup \{0\}$. The proposed method consists of two sub-procedures, as follows:

1. Classical computer preprocessing—Given the input of vector A , the method creates $f_A = (f_1, \dots, f_k)$ where each $f_i \in f_A$ represents the number of occurrences of the i^{th} item for $k \leq n$. Next, it transforms f_A to an amplitude vector, such that each $f_i \in f_A$ is converted to $\frac{\sqrt{f_i}}{\sqrt{\sum_{i=1}^k f_i}}$ and returns the normalized vector.
2. Quantum variance estimation—According to the input of the classical computer preprocessing results, the proposed method creates a quantum circuit and allocates $\lceil \log(k) \rceil + 1$ qubits. Once the circuit is ready, it initializes the state of the system by the normalized amplitude vector, denoted as $|\psi\rangle$, and the additional state vector, denoted as $|\psi'\rangle$, as follows²:

$$|\psi\rangle = \frac{1}{\sqrt{\sum f_A}} \sum_{j=1}^i \sqrt{f_j} |j\rangle$$

$$|\psi'\rangle = \frac{1}{\sqrt{\sum \left| \sqrt{j^{2r-ri}} \sqrt{f_j} \right|}} \sum_{j=1}^k \sqrt{j^{2r-ri}} \sqrt{f_j} |j\rangle$$

Next, the method applies the Hadamard gate to move the state into superposition. Thus, the current state can be presented as:

$$H^{\otimes k} |\psi\rangle = \frac{\sum_i (-1)^{\psi \cdot i} |i\rangle}{\sqrt{2^k}}$$

Once the state is in superposition, the method uses the parametrized gate (described in Section 2.1) to estimate the expected value of A and A^2 , denoted as φ_1, φ_2 , respectively. Thus, it calculates φ_1 as $\langle \psi' | M(k, 1) | \psi \rangle$ and φ_2 as $\langle \psi' | M(k, 2) | \psi \rangle$. Last, the method returns the value of $\varphi_2 - \varphi_1^2$ using simple classical computer computation. Figure 1

2.3 Correctness

Let $A = (a_1, a_2, \dots, a_n)$ be the input vector, and let $f_A = (f_1, \dots, f_k)$ be the occurrences of each item in A (i.e., each $f_i \in f_A$ represents the number of occurrences of the i^{th} item for $k \leq n$). The method transforms f_A to an amplitude vector as $|\psi\rangle = \frac{1}{\sqrt{\sum f_A}} \sum_{i=1}^k \sqrt{f_i} |i\rangle$ that satisfies:

$$\| |\psi\rangle \|^2 = \sum_{i=1}^k \left(\frac{\sqrt{f_i}}{\sqrt{\sum_{j=1}^k f_j}} \right)^2 = \sum_{i=1}^k \frac{f_i}{\sum_{j=1}^k f_j} = \frac{\sum_{i=1}^k f_i}{\sum_{j=1}^k f_j} = 1$$

² The parameter r is the same as in the parametrized gate.

Since a quantum system of k qubits provides 2^k amplitudes, encoding f_A requires the use of $\lceil \log_2 k \rceil + 1$ qubits. Notably, in cases where the length of f_A is not to the power of two, zeros are added as their values do not change the calculation. Given the initialized state vector, $|\psi\rangle$, The method applies the U gate with the parameters $\theta = \frac{\pi}{2}$, $\phi = 0$, $\lambda = \pi$, which is equivalent to applying the Hadamard gate to move the state into superposition (Wijesekera et al. 2009):

$$H^{\otimes k} |\psi\rangle = \frac{\sum_i (-1)^{w \cdot i} |i\rangle}{\sqrt{2^k}}$$

Next, the algorithm creates $M(a, r)$, a parameterized and diagonal square matrix of size a as described in Section 2.1. The algorithm uses $M(k, 1)$ and $M(k, 2)$ to estimate the values of the first and second moments of A . Let φ_1 be the expected value of applying $M(k, 1)$ on $|\psi\rangle$, denoted as $\langle \psi | M(k, 1) | \psi \rangle$ (Bakshi and Mahanthappa 1963). Given that $|\psi\rangle = \frac{1}{\sqrt{\sum f_A}} \sum_{i=1}^k \sqrt{f_i} |i\rangle$, the following proof of correctness describes a system with two qubits, although it can be generalized using the tensor product:

$$\begin{aligned} |\psi\rangle &= \frac{1}{\sqrt{\sum_{j=1}^k f_j}} \sum_{i=1}^k \sqrt{f_i} |i\rangle = \frac{\sqrt{f_1}}{\sqrt{\sum_{j=1}^4 f_j}} |00\rangle \\ &+ \frac{\sqrt{f_2}}{\sqrt{\sum_{j=1}^4 f_j}} |01\rangle + \frac{\sqrt{f_3}}{\sqrt{\sum_{j=1}^4 f_j}} |10\rangle + \frac{\sqrt{f_4}}{\sqrt{\sum_{j=1}^4 f_j}} |11\rangle \end{aligned}$$

$$\begin{aligned} \langle M(k, 1) | \psi \rangle &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} \begin{bmatrix} \sqrt{1^i} & 0 & 0 & 0 \\ 0 & \sqrt{2^i} & 0 & 0 \\ 0 & 0 & \sqrt{3^i} & 0 \\ 0 & 0 & 0 & \sqrt{4^i} \end{bmatrix} \begin{pmatrix} \sqrt{f_1} \\ \sqrt{f_2} \\ \sqrt{f_3} \\ \sqrt{f_4} \end{pmatrix} \\ &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} \begin{pmatrix} \sqrt{1^i} \sqrt{f_1} \\ \sqrt{2^i} \sqrt{f_2} \\ \sqrt{3^i} \sqrt{f_3} \\ \sqrt{4^i} \sqrt{f_4} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \langle \psi | M(k, 1) | \psi \rangle &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} \cdot \begin{pmatrix} \sqrt{1^{21-i}} \sqrt{f_1} \\ \sqrt{2^{21-i}} \sqrt{f_2} \\ \sqrt{3^{21-i}} \sqrt{f_3} \\ \sqrt{4^{21-i}} \sqrt{f_4} \end{pmatrix}^T \begin{pmatrix} \sqrt{1^i} \sqrt{f_1} \\ \sqrt{2^i} \sqrt{f_2} \\ \sqrt{3^i} \sqrt{f_3} \\ \sqrt{4^i} \sqrt{f_4} \end{pmatrix} \\ &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} (f_1 + 2f_2 + 3f_3 + 4f_4) \end{aligned}$$

It is important to note that since the norm of a complex number is a real number, it can easily be normalized to

facilitate the sum of squared amplitudes to equal one. Given that each f_i represents the normalized frequency of the i^{th} item, the expected value of the operator is equal to the first moment of A , i.e., its expectation. Similarly, let φ_2 be the predicted value of applying $M(k, 2)$ on $|\psi\rangle$, denoted as $\langle \psi | M(k, 2) | \psi \rangle$:

$$\begin{aligned} \langle M(k, 2) | \psi \rangle &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} \begin{bmatrix} \sqrt{1^{2i}} & 0 & 0 & 0 \\ 0 & \sqrt{2^{2i}} & 0 & 0 \\ 0 & 0 & \sqrt{3^{2i}} & 0 \\ 0 & 0 & 0 & \sqrt{4^{2i}} \end{bmatrix} \begin{pmatrix} \sqrt{f_1} \\ \sqrt{f_2} \\ \sqrt{f_3} \\ \sqrt{f_4} \end{pmatrix} \\ &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} \begin{pmatrix} \sqrt{1^{2i}} \sqrt{f_1} \\ \sqrt{2^{2i}} \sqrt{f_2} \\ \sqrt{3^{2i}} \sqrt{f_3} \\ \sqrt{4^{2i}} \sqrt{f_4} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \langle \psi | M(k, 2) | \psi \rangle &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} \begin{pmatrix} \sqrt{1^{2-2i}} \sqrt{f_1} \\ \sqrt{2^{2-2i}} \sqrt{f_2} \\ \sqrt{3^{2-2i}} \sqrt{f_3} \\ \sqrt{4^{2-2i}} \sqrt{f_4} \end{pmatrix}^T \begin{pmatrix} \sqrt{1^{2i}} \sqrt{f_1} \\ \sqrt{2^{2i}} \sqrt{f_2} \\ \sqrt{3^{2i}} \sqrt{f_3} \\ \sqrt{4^{2i}} \sqrt{f_4} \end{pmatrix} \\ &= \frac{1}{\sqrt{\sum_{j=1}^4 f_j}} (f_1 + 4f_2 + 9f_3 + 16f_4) \end{aligned}$$

Thus, the value of $\varphi_2 - \varphi_1^2$ is equal to $\mathbb{E}[A^2] - (\mathbb{E}[A])^2$ and represents its variance.³

3 Case Study

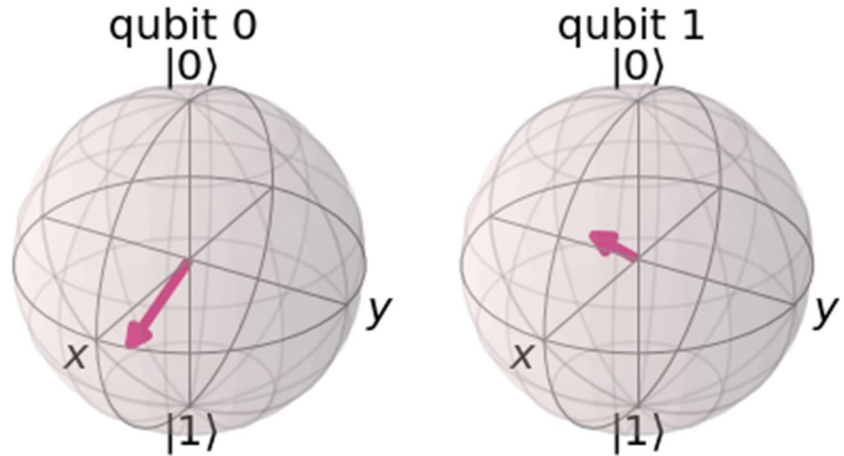
This section describes a case study and tests comparing the proposed method to a classical computer calculation. An IBM simulator with 1024 shots was used to simulate the trials. To simplify the illustration, a basic use case is first described regarding calculating the variance of a simple occurrences vector and details each operation in the quantum circuit. Then, the results of the proposed method are presented for the features of the diabetes dataset (Kahn 1994).

3.1 Simple occurrences vector

Let $A = (1, 2, 2, 2, 1, 4, 1, 4)$ be the input vector and let $f_A = (3, 3, 0, 2)$ be the occurrences vector of size four, such that the first item appeared three times, the second item

³ Since the norm of a complex number is a real number, the coefficients $|\psi\rangle$ can be normalized and the final output has the approximation value of $(\sum_{j=1}^k f_j)$, which can be handled using simple processing or multiplication using a classical computer.

Fig. 2 Initialized qubits in Bloch sphere representation



appeared three times, and so on. Since f_A has four values, it requires two qubits, and the initialized state is:

$$|\psi\rangle = \frac{1}{\sqrt{\sum_{j=1}^k f_j}} \sum_{i=1}^k \sqrt{f_i} |i\rangle$$

$$= \frac{1}{\sqrt{8}} (\sqrt{3}|00\rangle + \sqrt{3}|01\rangle + 0|10\rangle + \sqrt{2}|11\rangle)$$

Figure 2 presents the initialized state, $|\psi\rangle$, over two qubits in a Bloch sphere⁴ representation. Using the initialized state vector, the method used the $M(a, r)$ parametrized gate and estimated its expected value in two manners:

$$\langle M(k, 1) | \psi \rangle = \frac{1}{\sqrt{8}} \begin{bmatrix} \sqrt{1^i} & 0 & 0 & 0 \\ 0 & \sqrt{2^i} & 0 & 0 \\ 0 & 0 & \sqrt{3^i} & 0 \\ 0 & 0 & 0 & \sqrt{4^i} \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ \sqrt{3} \\ \sqrt{0} \\ \sqrt{2} \end{bmatrix}$$

$$= \frac{1}{\sqrt{8}} \begin{pmatrix} \sqrt{1^i} \sqrt{3} \\ \sqrt{2^i} \sqrt{3} \\ 0 \\ \sqrt{4^i} \sqrt{2} \end{pmatrix}$$

$$\varphi_1 = \langle \psi | M(k, 1) | \psi \rangle = \frac{1}{\sqrt{8}} \begin{pmatrix} \sqrt{1^{2-2i}} & \sqrt{3} \\ \sqrt{2^{2-2i}} & \sqrt{3} \\ \sqrt{3^{2-2i}} & \sqrt{0} \\ \sqrt{4^{2-2i}} & \sqrt{2} \end{pmatrix}^T \begin{pmatrix} \sqrt{1^i} \sqrt{3} \\ \sqrt{2^i} \sqrt{3} \\ 0 \\ \sqrt{4^i} \sqrt{2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{8}} (1 \cdot 3 + 2 \cdot 3 + 0 + 4 \cdot 2) = \frac{2.125}{\sqrt{8}}$$

$$\langle M(k, 2) | \psi \rangle = \frac{1}{\sqrt{8}} \begin{bmatrix} \sqrt{1^i} & 0 & 0 & 0 \\ 0 & \sqrt{2^i} & 0 & 0 \\ 0 & 0 & \sqrt{3^i} & 0 \\ 0 & 0 & 0 & \sqrt{4^i} \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ \sqrt{3} \\ \sqrt{0} \\ \sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{8}} \begin{pmatrix} \sqrt{1^i} \sqrt{3} \\ \sqrt{2^i} \sqrt{3} \\ 0 \\ \sqrt{4^i} \sqrt{2} \end{pmatrix}$$

$$\varphi_2 = \langle \psi | M(k, 2) | \psi \rangle = \frac{1}{\sqrt{8}} \begin{pmatrix} \sqrt{1^{2-2i}} & \sqrt{3} \\ \sqrt{2^{2-2i}} & \sqrt{3} \\ \sqrt{3^{2-2i}} & \sqrt{0} \\ \sqrt{4^{2-2i}} & \sqrt{2} \end{pmatrix}^T \begin{pmatrix} \sqrt{1^i} \sqrt{3} \\ \sqrt{2^i} \sqrt{3} \\ 0 \\ \sqrt{4^i} \sqrt{2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{8}} (3 + 2^2 \cdot 3 + 0 + 4^2 \cdot 2) = \frac{5.875}{\sqrt{8}}$$

Lastly, using the classical computer computation, the method returned:

$$\varphi_2 - \varphi_1^2 = \frac{5.875}{\sqrt{8}} - \frac{2.125^2}{8} = 1.512$$

The classical computer calculation yielded a variance of 1.553, and the error between the classical and quantum computation was 0.021.

3.2 The diabetes dataset

In this section, using a dataset, the proposed method is compared with the classical computer method and analyzed. For the comparison, we used the diabetes dataset (Kahn 1994), which includes 768 diabetic and non-diabetic women. It consists of eight features and a Boolean target variable. The ‘‘BMI’’ and ‘‘DiabetesPedigreeFunction’’ features were removed since the proposed method is designed for discrete values. A classical computer was used to calculate each feature’s variance, and an IBM simulator and the proposed method were used to

⁴ A quantum state can be presented in a Bloch sphere as $\cos(\frac{\theta}{2})|0\rangle + e^{i\phi} \sin(\frac{\theta}{2})|1\rangle$ for $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$.

calculate its quantum variance. Table 1 compares the variance values achieved for each feature and presents the error rate.

The results showed the consistency of the proposed method across six features of the data set. The minimal error occurred in the “Age” column with a deviation of only 0.021% of the variance, calculated using a classic computer. The maximum error was obtained in the “Blood Pressure” column with an error of 5.996%. The calculation of the distribution measures of each feature in the dataset according to the proposed method was consistent and showed a high agreement with the original value. At the same time, further analysis is required to examine the proposed method, as will be presented in Section 4.

To understand the behavior of the “Blood Pressure” feature, which raised a maximum error of 5.996%, Fig. 3 shows its distribution estimation. It is known that anomalous values in the distribution can cause biased results and a wide error range when calculating distribution measures, such as mean and variance. Approximately 40 records were defined as outliers, which may have caused a significant change in the variance estimation.

4 Results

Five discrete value distributions were compared to assess and evaluate the results of the proposed method (Table 2). For each of the following distributions, 10,000 experiments were conducted:

1. Binomial distribution, $Bin(n, p)$, with a success probability of p in a total of n trials.
2. Geometric distribution, $G(p)$, with success probability p .
3. Uniform distribution between $a, b \in \mathbb{N}$, denoted as $U(a, b)$.
4. Hypergeometric distribution, $HG(N, D, n)$, with a total of N items, D specials, and n trials.
5. Poisson distribution, $Pois(\lambda)$, where λ is the expected value of events in an interval of time.

Table 1 A variance comparison of the diabetes dataset features

	Variance	Quantum variance	Error
Pregnancies	11.354	11.002	0.352 (3.100%)
Glucose	1022.248	1005.813	16.435 (1.607%)
Blood Pressure	374.647	397.111	22.464 (5.996%)
Skin Thickness	254.473	243.556	10.917 (4.290%)
Insulin	13,281.180	13,496.174	214.994 (1.618%)
Age	138.303	138.333	0.030 (0.021%)

The “Error” column is presented as A(B), where A is the error rate and B is the percentage of change between the values

The results of the comparison between the distributions showed a stronger consistency than the results of the diabetes dataset (described in Section 3.2). The highest error had a value of 1.351, which occurred in the binomial distribution with a probability of success of 0.7 in a single trial. For all the examined distributions, the quantum method maintained low error values, which reinforced the implementation of the proposed method and its results.

To examine the effect of the input size (the number of qubits) on the performance of our method, the following experiments were conducted:

1. Binomial distribution—Inputs of sizes 2^i for $3 \leq i \leq 8$ were created, which required i qubits, respectively. The binomial distribution was used due to the ability to control the input size (unlike geometric distribution). The variance obtained was compared using a classical and quantum calculation for each of the probabilities between 0.1 and 0.9 with a step of 0.1 (i.e., for each examined input, there was a total of nine pairs of variances). Figure 4 presents this comparison, where each dot in the figure represents a probability. For example, nine qubits have a total of 18 points (nine for the quantum calculation and nine for the classic calculation). Figure 4 examines the level of agreement between the quantum method and the classical calculation for different variance values. In cases without agreement between the quantum and classical calculation, Fig. 4 would present a grouping of red and black points separately.
2. Hypergeometric distribution—Inputs of size 10,000 were created, with a total of $N = 500$, $n = 200$, and the values of 2^i were examined, where $2 \leq i \leq 7$ for D . Figure 5 shows the comparison between the calculated variance over different values of D . Like Fig. 4, the x -axis represents the number of qubits required to encode the input, and the y -axis shows the variances obtained for different D values.

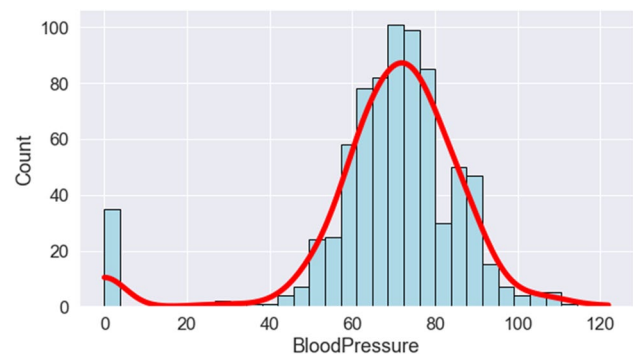


Fig. 3 The distribution of the “Blood Pressure” feature

Table 2 Variance comparison over discrete value distributions

Distribution	Parameters	Variance	Quantum variance	Error
$Bin(n, p)$	$p = 0.1$	90.793	91.874	1.080
	$p = 0.3$	208.859	210.200	1.341
	$p = 0.5$	249.617	250.252	0.634
	$p = 0.7$	209.990	211.341	1.351
	$p = 0.9$	91.320	92.623	1.302
$G(p)$	$p = 0.1$	94.463	93.191	1.272
	$p = 0.3$	7.831	6.956	0.874
	$p = 0.5$	2.010	1.525	0.484
	$p = 0.7$	0.622	0.905	0.282
	$p = 0.9$	0.127	1.132	1.005
$U(a, b)$	$a = 1, b = 10$	8.230	8.811	0.581
$HG(N, D, n)$	$N = 55, D = 2, n = 32$	0.424	0.612	0.188
$Pois(\lambda)$	$\lambda = 10$	10.072	10.991	0.919

According to the results of the binomial distribution, it can be concluded that the proposed method showed reliable performance, even among data with high variability (i.e., high number of qubits). This can be seen in the agreement between the red points (which represent quantum calculation) and the gray points (which represent classical calculation). Similar to the binomial distribution, there was consistency among the hypergeometric distribution between

the calculation of the reefs. However, small inputs (which required 2 or 3 qubits) showed a more comprehensive range of answers and gaps between the classical and quantum computations. These were not gaps that presented an error; rather, they illustrated the difficulty in estimating the exact values on a quantum computer.

5 Conclusions and discussion

This study proposes a novel quantum-parametrized gate and circuit implementation for distribution measure estimations. The presented method can be implemented in data analysis processes, machine learning techniques, and more. Its main innovation is the use of a parametrized quantum circuit to calculate the expectation and variance of a given vector. As a result, this method is accessible to those without a previous understanding of QC. The following are the main conclusions:

1. The parametrized quantum gate proposed in this study was found to be effective in estimating the distribution values (i.e., expectation and variance) of a discrete value vector. When comparing the proposed and classical methods using the diabetes dataset (Section 3.2), the error range of the obtained variance ranged between

Fig. 4 A comparison between the binomial variance over different numbers of qubits

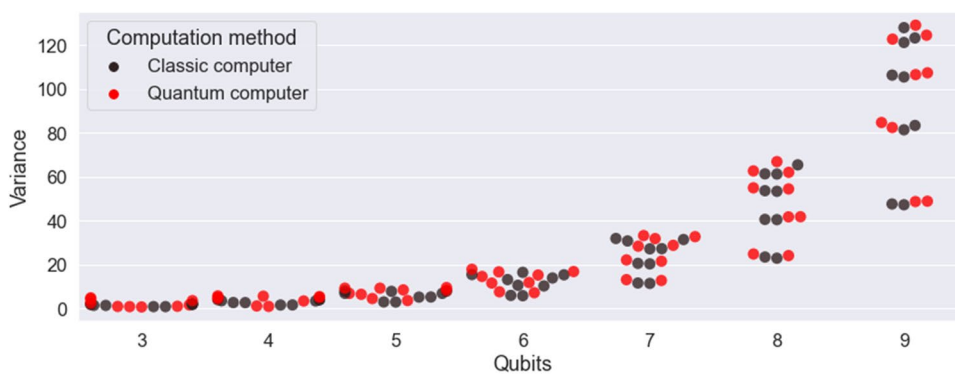
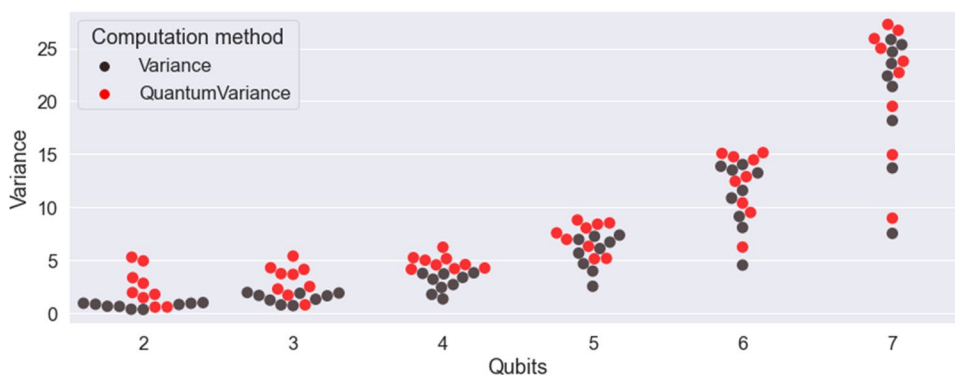


Fig. 5 A comparison between the hypergeometric variance over different numbers of qubits



0.021% and 5.996%. The feature for which the maximum error was obtained, “Blood Pressure,” was a noisy column, and therefore, further research is required.

- In testing the discrete distributions (Section 4), inputs of different sizes were compared to examine the effect of the number of required qubits and the obtained result. A wide agreement was found between the variance calculation using a classical computer and the proposed method in QC. In these cases, in contrast to the data set presented in Section 3.2, a minimal error was obtained, even for noisy distributions with inconsistent variance values.

The limitation of this study is expressed in the constraint of the method only being able to use discrete values due to the classical computer preprocessing process, which includes the creation of a frequency vector. Nevertheless, this study presented four main issues that should be addressed in future studies. First, the generalization of the parametric gate and its adaptation to the need to calculate additional distribution measures, such as skewness, kurtosis, and more, should be further examined. Second, a deep study should be conducted on the parameter values presented in this study and their optimal values for minimizing the error value. Third, an adaptation of the method for continuous distributions such as normal, exponential, and more would be beneficial. Last, using amplitude encoding may be efficient for a logarithmic number of qubits, although it requires a large number of quantum gates. Future studies are encouraged to explore the relationship between the input encoding and the number of quantum gates required to represent it to optimize the trade between these variables.

Author contributions The authors confirm contribution to the paper as follows: conceptualization: OP; methodology: OP and MK; data collection: OP and MK; data analysis: MK; draft preparation: OP and MK; writing and review: MK; editing: OP. All authors reviewed the results and approved the final version of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Shenkar College of Engineering and Design.

Data availability The data used in this study has been cited and is available and accessible in UCI Machine Learning Repository – Datasets (<https://archive.ics.uci.edu/>).

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alchieri L, Badalotti D, Bonardi P, Bianco S (2021) An introduction to quantum machine learning: from quantum logic to quantum deep learning. *Quantum Mach Intell* 3:28. <https://doi.org/10.1007/s42484-021-00056-8>
- Assouel A, Jacquier A, Kondratyev A (2022) A quantum generative adversarial network for distributions. *Quantum Mach Intell* 4:28. <https://doi.org/10.1007/s42484-022-00083-z>
- Azevedo V, Silva C, Dutra I (2022) Quantum transfer learning for breast cancer detection. *Quantum Mach Intell* 4:5. <https://doi.org/10.1007/s42484-022-00062-4>
- Bakshi PM, Mahanthappa KT (1963) Expectation value formalism in quantum field theory. *I J Math Phys* 4(1):1–11. <https://doi.org/10.1063/1.1703883>
- Behrens JT, Yu CH (2003) Exploratory data analysis. In: Schinka JA, Velicer WF (eds) *Handbook of psychology*. John Wiley & Sons, New Jersey, pp 33–64
- Benedetti M, Lloyd E, Sack S, Fiorentini M (2019) Parameterized quantum circuits as machine learning models. *Quantum Sci Technol* 4:043001. <https://doi.org/10.1088/2058-9565/ab4eb5>
- Bennett CH, Bernstein E, Brassard G, Vazirani U (1997) Strengths and weaknesses of quantum computing. *SIAM J Comput* 26:1510–1523. <https://doi.org/10.1137/S0097539796300933>
- Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. *Nature* 549:195–202. <https://doi.org/10.1038/nature23474>
- Boyer M, Brassard G, Høyer P, Tapp A (1998) Tight bounds on quantum searching. *Fortschritte Der Phys* 46:493–505. [https://doi.org/10.1002/\(SICI\)1521-3978\(199806\)46:4<5%3c493::AID-PROP493%3e3.0.CO;2-P](https://doi.org/10.1002/(SICI)1521-3978(199806)46:4<5%3c493::AID-PROP493%3e3.0.CO;2-P)
- Bravo-Prieto C (2021) Quantum autoencoders with enhanced data encoding. *Mach Learn: Sci Technol* 2:035028. <https://doi.org/10.1088/2632-2153/ac0616>
- Buffoni L, Caruso F (2021) New trends in quantum machine learning (a). *Europhys Lett* 132:60004. <https://doi.org/10.1209/0295-5075/132/60004>
- Chatfield C (1986) Exploratory data analysis. *Eur J Oper Res* 23:5–13. [https://doi.org/10.1016/0377-2217\(86\)90209-2](https://doi.org/10.1016/0377-2217(86)90209-2)
- Cleff T (2014) Exploratory data analysis in business and economics. Springer, Cham <https://doi.org/10.1007/978-3-319-01517-0>
- Cross A (2018) The IBM Q experience and QISKit open-source quantum computing software. *APS March Meet Abstr* 2018:L58-003
- DallaPozza N, Buffoni L, Martina S, Caruso F (2022) Quantum reinforcement learning: The maze problem. *Quantum Mach Intell* 4:11. <https://doi.org/10.1007/s42484-022-00068-y>
- Dilip R, Liu YJ, Smith A, Pollmann F (2022) Data compression for quantum machine learning. *Phys Rev Res* 4:043007. <https://doi.org/10.1103/PhysRevResearch.4.043007>
- Du Y, Hsieh MH, Liu T, Tao D (2020) Expressive power of parametrized quantum circuits. *Phys Rev Res* 2:033125. <https://doi.org/10.1103/PhysRevResearch.2.033125>
- Gelman A (2004) Exploratory data analysis for complex models. *J Comput Graph Stat* 13:755–779. <https://doi.org/10.1198/106186004X11435>

- González FA, Gallego A, Toledo-Cortés S, Vargas-Calderón V (2022) Learning with density matrices and random features. *Quantum Mach Intell* 4:23. <https://doi.org/10.1007/s42484-022-00079-9>
- Hubregtsen T, Pichlmeier J, Stecher P, Bertels K (2021) Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility, and entangling capability. *Quantum Mach Intell* 3:1–19
- Hur T, Kim L, Park DK (2022) Quantum convolutional neural network for classical data classification. *Quantum Mach Intell* 4:3. <https://doi.org/10.1007/s42484-021-00061-x>
- Jebb AT, Parrigon S, Woo SE (2017) Exploratory data analysis as a foundation of inductive research. *Hum Resour Manag Rev* 27:265–276. <https://doi.org/10.1016/j.hrmr.2016.08.003>
- Kahn M (1994) Diabetes. *UCI Mach Learn Repositor* <https://doi.org/10.24432/C5T59G>
- Komorowski M, Marshall DC, Saliccioli JD, Crutain Y (2016) Exploratory data analysis. In: MIT Critical Data (ed) Secondary analysis of electronic health records. Springer, Cham, pp 185–203.
- Koren M, Koren O, Peretz O (2023) A quantum “black box” for entropy calculation. *Quantum Mach Intell* 5:37. <https://doi.org/10.1007/s42484-023-00127-y>
- LaRose R, Coyle B (2020) Robust data encodings for quantum classifiers. *Phys Rev A* 102:032420. <https://doi.org/10.1103/PhysRevA.102.032420>
- Leinhardt S, Wasserman SS (1979) Exploratory data analysis: an introduction to selected methods. *Sociol Methodol* 10:311–365. <https://doi.org/10.2307/270776>
- Li G, Ye R, Zhao X, Wang X (2022) Concentration of data encoding in parameterized quantum circuits. *Adv Neural Inf Process Syst* 35:19456–19469
- Majji SR, Chalumuri A, Manoj BS (2023) Quantum approach to image data encoding and compression. *IEEE Sens Lett* 7:1–4. <https://doi.org/10.1109/LSENS.2023.3239749>
- Morgenthaler S (2009) Exploratory data analysis. *Wiley Interdiscip Rev Comput Stat* 1:33–44. <https://doi.org/10.1002/wics.2>
- Peham T, Burgholzer L, Wille R (2023). Equivalence checking of parameterized quantum circuits: verifying the compilation of variational quantum algorithms. In: Proceedings of the 28th Asia and South Pacific Design Automation Conference, January 2023. Association for Computing Machinery, New York, pp 702–708. <https://doi.org/10.1145/3566097.3567932>
- Piattini M, Peterssen G, Pérez-Castillo R (2021) Quantum computing: a new software engineering golden age. *ACM SIGSOFT Software Engineering Notes* 45:12–14
- Pirhooshyaran M, Terlaky T (2021) Quantum circuit design search. *Quantum Mach Intell* 3:25. <https://doi.org/10.1007/s42484-021-00051-z>
- Robertson JK (1943) The role of physical optics in research. *Am J Phys* 11:264–271. <https://doi.org/10.1119/1.1990496>
- Romero J, Olson JP, Aspuru-Guzik A (2017) Quantum autoencoders for efficient compression of quantum data. *Quantum Sci Technol* 2:045001. <https://doi.org/10.1088/2058-9565/aa8072>
- Shin S, Teo YS, Jeong H (2023) Exponential data encoding for quantum supervised learning. *Phys Rev A* 107:012422. <https://doi.org/10.1103/PhysRevA.107.012422>
- Sierra-Sosa D, Pal S, Telahun M (2023) Data rotation and its influence on quantum encoding. *Quantum Inf Process* 22:89. <https://doi.org/10.1007/s11128-023-03837-1>
- Tukey JW (1977) Exploratory data analysis, vol 2. Addison-Wesley, Reading, MA
- Tüysüz C, Rieger C, Novotny K, Demirköz B, Dobos D, Potamianos K, Vallecorsa S, Vilmant JR, Forster R (2021) Hybrid quantum classical graph neural networks for particle track reconstruction. *Quantum Mach Intell* 3:29. <https://doi.org/10.1007/s42484-021-00055-9>
- Vigni ML, Durante C, Cocchi M (2013) Exploratory data analysis. In: Marini F (ed) Data handling in science and technology, vol 28. Elsevier, pp 55–126. <https://doi.org/10.1016/B978-0-444-59528-7.00003-X>
- Wang G, Zhao B, Wu B, Zhang C, Liu W (2023) Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases. *Int J Min Sci Technol* 33:47–59. <https://doi.org/10.1016/j.ijmst.2022.07.002>
- Weigold M, Barzen J, Leymann F, Salm M (2021a) Expanding data encoding patterns for quantum algorithms. In: 2021 IEEE 18th International Conference on Software Architecture Companion (ICSA-C). Stuttgart, Germany, pp 95–101. <https://doi.org/10.1109/ICSA-C52384.2021.00025>
- Weigold M, Barzen J, Leymann F, Salm M (2021b) Encoding patterns for quantum algorithms. *IET Quantum Commun* 2:141–152. <https://doi.org/10.1049/qt2.12032>
- Wiebe N (2020) Key questions for the quantum machine learner to ask themselves. *New J Phys* 22:091001. <https://doi.org/10.1088/1367-2630/abac39>
- Wijesekera S, Huang X, Sharma D (2009) Multi-agent based approach for quantum key distribution in WiFi networks. Agent and multi-agent systems: technologies and applications: third KES International Symposium, KES-AMSTA 2009, Uppsala, Sweden, June 3–5, 2009. Springer, Berlin Heidelberg, pp 293–303
- Wongsuphasawat K, Liu Y, Heer J (2019) Goals, process, and challenges of exploratory data analysis: an interview study. *arXiv:1911.00568*. <https://doi.org/10.48550/arXiv.1911.00568>
- Ying M (2010) Quantum computation, quantum theory and AI. *Artif Intell* 174:162–176. <https://doi.org/10.1016/j.artint.2009.11.009>
- Zeng W, Johnson B, Smith R, Rubin N, Reagor M, Ryan C, Rigetti C (2017) First quantum computers need smart software. *Nature* 549:149–151. <https://doi.org/10.1038/549149a>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.