



Rapid communication

Likelihood theory in a quantum world: Tests with quantum coins and computers

ARPITA MAITRA^{1,*}, JOSEPH SAMUEL^{2,3} and SUPURNA SINHA²

¹C.R. Rao Advanced Institute of Mathematics, Statistics and Computer Science, University of Hyderabad Campus, Prof. C.R. Rao Road, Gachibowli, Hyderabad 500 046, India

²Raman Research Institute, Sadashivanagar, Near Mekhri Circle, Bengaluru 560 080, India

³International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Shivakote, Bengaluru 560 089, India

*Corresponding author. E-mail: arpita76b@gmail.com

MS received 20 September 2019; revised 19 November 2019; accepted 24 December 2019

Abstract. By repeated trials, one can determine the fairness of a classical coin with a confidence which grows with the number of trials. A quantum coin can be in a superposition of heads and tails and its state is most generally a density matrix. Given a string of qubits representing a series of trials, one can measure them individually and determine the state with a certain confidence. We show that there is an improved strategy which measures the qubits after entangling them, which leads to a greater confidence. This strategy is demonstrated on the simulation facility of IBM quantum computers.

Keywords. Likelihood theory; quantum state discrimination; IBM quantum computer; quantum advantage.

PACS Nos 03.67.–a; 03.67.Lx; 03.65.Ud

1. Introduction

When testing a theory against experimental data, the Bayesian approach gives us a rational way of revising our theoretical expectations in the light of new data. To take the simple and familiar example of coin tossing, let us start with the belief that our coin is fair. If we then toss the coin ten times and turn up nine heads, our belief in fairness will be shaken but not destroyed: it is still possible that the nine heads were generated by chance. If the run of heads continues, we would be hard pressed to cling to our belief in the fairness of the coin. Revising our beliefs in the light of new data is an essential component of the scientific method, a point which was strikingly brought out in the Monty Hall problem, which occupied the community in the 1990's. Bayesian theory is routinely used in testing the efficacy of drugs, where one starts with the 'null hypothesis' that the drug being tested is no more effective than a placebo. If the

drug fares better than a placebo over a sufficiently large number of trials, we are more inclined to believe in its efficacy.

Let us now return to coin tossing, which captures the idea of Bayesian inference in its simplest form. The intuitive idea is quantitatively captured in likelihood theory. Our initial belief (prior) is that the probability of heads is q_H and tails is q_T . If we see a particular string $S_N = \{HTHHHHTTHTTT\dots\}$, in N coin tosses, the likelihood of such a string emerging from the distribution $P_B = \{q_H, q_T\}$ is given by (see below for a derivation)

$$L(S_N|P_B) = \frac{1}{\sqrt{2\pi N p_H p_T}} \exp -\{ND_{KL}(P_A||P_B)\}, \quad (1)$$

where $P_A = \{p_H, p_T\} = \{N_H/N, N_T/N\}$ is the observed frequency distribution in a string of N tosses. $D_{KL}(P_A||P_B)$ is the relative entropy or the Kullback–Leibler divergence between the distributions P_B and P_A .

Electronic supplementary material: The online version of this article (<https://doi.org/10.1007/s12043-020-1926-9>) contains supplementary material, which is available to authorized users.

When we pass from classical coins to quantum coins, a new possibility emerges: the coins may be in a superposition of heads $|H\rangle$ and tails $|T\rangle$. Even more generally, the state of the system need not even be pure but a density matrix. The probability distributions P_B and P_A describing classical coins are replaced by density matrices ρ_B and ρ_A . One of the central problems in quantum information is quantum state discrimination. Given two quantum states, how easily can we tell them apart? This question has led to the issue of distinguishability measures on the space of quantum states [1–5]. More recently, in [6] it has been noticed that quantum entanglement leads to a significant improvement in distinguishing two quantum states. In this paper, we go beyond [6] by actually implementing our theoretical model on a state-of-the-art quantum computer, the IBM quantum computer and checking our theoretical ideas against quantum simulations and experiments.

The problem of quantum state discrimination is explained in a review by Helstrom [7]. Given a string of data consisting entirely of either ρ_A or ρ_B states and additional information such as prior probability of these two possibilities one has to make a determination of the state of the string. This version has been treated in ref. [8] which arrives at a quantum version of the Chernoff bound. The present work addresses a related but distinct problem. We do not assume any prior knowledge of the relative probabilities of the two states. We are concerned with the optimal measurement to be made on the data so as to maximise our ability to distinguish the two states. In this respect, the problem is a quantum version of classical likelihood theory. Given a string of data, how likely is it that it is described by a quantum state ρ_B . The challenge is to find the optimal measurement basis. Given two general states ρ_A and ρ_B , there does not exist a general method for finding the optimal measurement that will yield the best discrimination. In the limit that the two states are infinitesimally separated, Helstrom [7,9] elegantly answers this question by using the symmetric logarithmic derivative. In general, one has to use numerical optimisation, which is the approach we take here.

Our goal in this paper is to describe an experiment which demonstrates the use of entanglement as a resource in state discrimination [10]. We have successfully implemented our ideas on the IBM quantum simulator. In experimental runs, it appears that the quantum advantage we seek to demonstrate is swamped by noise. The experiment may well be within the reach of more sophisticated quantum computers which are currently in use and so provides a short term goal possibly within reach.

The outline of this paper is as follows. We first review the classical likelihood theory to set the background

for our work. We then describe how likelihood theory is modified in a quantum world and show how these abstract ideas can be translated into reality using existing quantum computers. We discuss the simplest direct measurement strategy where one measures qubits one by one. We then show that there exist improved strategies which exploit quantum entanglement in an essential way. We translate our abstract ideas into quantum scores, combinations of quantum logic gates, which can be run on quantum computers. Finally, we present the results of running our programs on the existing IBM quantum computers and end with some concluding remarks.

2. Review of classical likelihood theory

Let us consider a biased coin for which the probability of getting a head is $p_H = 1/3$ and that of getting a tail is $p_T = 2/3$. Suppose we incorrectly assume that the coin is fair and assign probabilities $q_H = 1/2$ and $q_T = 1/2$ for getting a head and a tail respectively. The question of interest is the number of trials needed to be able to distinguish (at a given confidence level) between our assumed probability distribution and the measured probability distribution. A popular measure for distinguishing between the expected distribution and the measured distribution is given by the relative entropy or the Kulback–Leibler divergence (KLD) which is widely used in the context of distinguishing classical probability distributions [11]. Let us consider N independent tosses of a coin leading to a string $S_N = \{HTHHTHTHHTTTTT\dots\dots\}$. What is the probability that the string S is generated by the model distribution $P_B = \{q, 1 - q\}$? The observed frequency distribution is $P_A = \{p, 1 - p\}$. If there are N_H heads and N_T tails in the string S_N then the probability of getting such a string is $(N!/N_H!N_T!)q^{N_H}(1 - q)^{N_T}$ which we call the likelihood function $L(S_N|P_B)$. If we take the average of the logarithm of this likelihood function and use Stirling's approximation for large N we get the following expression:

$$\frac{1}{N} \log L(N|P_B) = -D_{KL}(P_A||P_B) + \frac{1}{N} \log \frac{1}{\sqrt{2\pi Np(1-p)}}, \quad (2)$$

where

$$p = \frac{N_H}{N}$$

and

$$D_{KL}(P_A||P_B) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}.$$

The second term in (2) is due to the subleading term $\frac{1}{2} \log 2\pi N$ of Stirling’s approximation. If $D_{KL}(P_A||P_B) \neq 0$ then the likelihood of the string S being produced by the P_B distribution decreases exponentially with N . This results in eq. (1). Thus, $D_{KL}(P_A||P_B)$ gives us the divergence of the measured distribution from the model distribution. KLD is positive and vanishes if and only if the two distributions P_A and P_B are equal.

In this limit, we find that the exponential divergence gives way to a power-law divergence, as shown by the subleading term in (2). The arguments above generalise appropriately to an arbitrary number of outcomes (instead of two) and also to continuous random variables.

3. Quantum likelihood theory

In passing from classical likelihood theory described above to the quantum one, there are two additional features to be considered. First, as mentioned earlier, because of the superposition principle, a quantum coin would in general be described by a density matrix rather than a probability distribution. Second, in measuring a quantum system, one has an additional freedom: the choice of measurement basis. It is evident that the choice of measurement basis can affect our discriminating power between competing quantum states. We would be well advised to choose the basis wisely, so as to maximise our discriminating power between two given density matrices.

Consider two mixed density matrices ρ_A and ρ_B . Suppose that the (prior) quantum distribution is described by ρ_B . For a choice $\{m_i\}$ ($i = 1, \dots, n$) of orthonormal basis, the probable outcomes have the probabilities $q_i = \langle m_i | \rho_B | m_i \rangle$. Similarly, if the density matrix is ρ_A , we have the probabilities $p_i = \langle m_i | \rho_A | m_i \rangle$. Both $\{q_i\}$ and $\{p_i\}$ are normalised probability distributions. These can now be regarded as classical probability distributions and we can compute their relative entropy, defined as $D_{KL}(m) = \sum_{i=1, \dots, n} p_i \log(p_i/q_i)$. The argument m on the LHS indicates that the relative entropy ‘depends’ on the measurement basis m , since the p s and q s do. Let us now choose $m = m^*$ optimally so as to maximise $D_{KL}(m)$ which is a measure of our discriminating power. The basis-optimised value of the KLD is denoted by $S(A||B) = D_{KL}(m^*)$. For example, we could be detecting a weak signal (as in quantum metrology) using a quantum detector, which would remain in the state ρ_B in the absence of an incoming signal, but get excited to state ρ_A when the signal is present. If the state is ρ_A , our confidence C that the state is in fact ρ_B will decrease at the rate

$$C = \frac{1}{\sqrt{2\pi N p(1-p)}} \exp\{-NS(A||B)\},$$

where N is the number of measured systems.

3.1 Tossing quantum coins

The abstract theory of the last section can be converted into concrete experiments using the recent progress in constructing quantum computers in the laboratory. Choosing the optimal basis is effected by a unitary transformation, which can be implemented using quantum gates. The quantum circuit that we implement on the quantum computer can be divided into three parts. The first part is the state preparation, the second is the unitary transformation and the third is the measurement. We explore two strategies, which we call the direct and the entangling strategies. The two strategies differ only in the second part. In the direct strategy, we measure the qubits one at a time, taking care to choose the measurement basis so as to maximise our ability to distinguish the two states. In the entangling strategy, we perform an entangling unitary transformation on a pair of qubits before performing the measurement.

State preparation: Without loss of generality, we can assume the state ρ_A to be along the z direction of the Bloch sphere and ρ_B to lie in the $x-z$ plane of the Bloch sphere, making an angle δ with the z -axis. To prepare the initial quantum state ρ_A , which is a density matrix, we begin with two qubits (say a, b) both in the state $|0\rangle: |0\rangle_a \otimes |0\rangle_b$. We then perform a local unitary rotation through an angle β in the $x-z$ plane on one of them (b) and arrive at $|0\rangle_a \otimes (\cos \beta/2 |0\rangle_b + \sin \beta/2 |1\rangle_b)$. We then apply a CNOT gate with b as the control and a as the target. The total system is now in an entangled state. If we ignore (trace over) the qubit a , we get an impure density matrix on qubit b : $\rho_A = (\cos \beta/2)^2 |0\rangle \langle 0| + (\sin \beta/2)^2 |1\rangle \langle 1|$. This is our initial state ρ_A , whose purity depends on the entangling angle β . In runs we have used $\beta = 0.2$ and $\delta = 1.8$.

In practice (and for easier comparison with the entangling strategy) we repeat the arrangement with two more qubits with qubits a and b replaced respectively by d and c . We get identical states in qubits b and c . The rest of the circuit performs a unitary transformation on the qubits b and c and then performs a measurement in the computational basis.

Direct strategy: In the direct measurement strategy we perform measurements on individual qubits. One can show [12] that the optimal basis consists of two antipodal points in the $x-z$ plane of the Bloch sphere. This determines a direction in the $x-z$ plane which is characterised by an angle φ in the range $0-\pi$. We optimise

over φ by choosing φ as the measurement basis so that $D_{KL}(\varphi)$ defined in the last section is as large as possible. Then

$$S_{\text{rel}} = D_{KL}(\varphi^*)$$

is the maximum value where φ^* is the value of φ corresponding to the maximum value of D_{KL} . This gives us the optimal one-qubit strategy for the distinguishability of ρ_A and ρ_B . This is our direct measurement strategy. It is easy to find numerically the optimal unitary transformation for the direct strategy. The unitary transformations of interest are in fact orthogonal [12], and belong to $SO(2)$ (because the states are assumed to be in the x - z plane) and characterised by a single angle φ . We numerically evaluate the relative entropy as a function of φ and find the maximum φ^* . We use this unitary transformation U_2 to get the optimised measurement basis, which is implemented as a local transformation on qubits b and c . The logic circuit which implements the direct strategy on a quantum computer is displayed in figure 1.

Entangling strategy: The entangling strategy for discriminating qubits gives us an advantage over the direct measurement strategy, in quantum state discrimination. Consider $2N$ independent and identically distributed qubits. Let us group these $2N$ qubits into N pairs. We now perform measurements on each pair, choosing our measurement basis to give us optimal results for state distinguishability. This is effected by performing a unitary transformation U_4 in the two-qubit Hilbert space followed by measurement in the computational basis. By grouping the qubits in pairs, we have increased the freedom in the choice of measurement basis because we can choose entangled bases as well as separable ones. This greater freedom means that we can improve on the direct strategy. Maximising S_{rel} over ‘all’ two qubit bases, is clearly an improvement over maximising over separable ones.

We maximise the relative entropy in the space of two qubit bases, by doing a numerical search. Computer experiments reveal that one does indeed gain by using the entangling strategy. Our initial aim was to show that this entangling strategy indeed gives us an improvement over the direct measurement strategy which we have described in the last section, by explicitly implementing it on a quantum computer.

For the entangled strategy, we find it advantageous to simplify our search by searching within $SO(4)$. We perform a random walk in the $SO(4)$ space to find the optimal U_4 . This yields a relative entropy per qubit which is higher than that obtained by the direct strategy.

Having determined a U_4 transformation, we now have to construct a quantum score that implements the

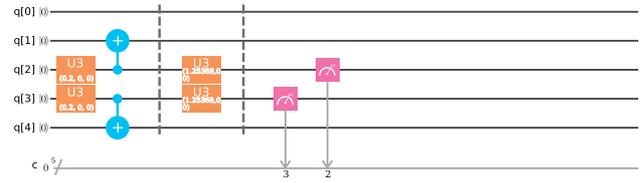


Figure 1. The quantum logic circuit for the direct strategy. The unitary transformation to the optimal separable basis is between the two black dashed vertical lines. The preparation stage is the part to the left of these lines and to the right of these lines, the pink blocks with dials show the measurement stage.

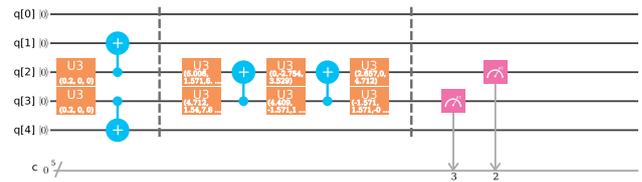


Figure 2. The quantum logic circuit for the entangling strategy. The unitary transformation to the optimal entangling basis is between the two black dashed vertical lines. This entangling unitary transformation is realised (upto an overall phase) as a string of unitary operations and two CNOT gates. The preparation stage is the part to the left of these lines and to the right of these lines, the pink blocks with dials show the measurement stage.

transformation. The basic units available are CNOT gates (which operate on two qubits) and local unitaries (which operate on a single qubit). As the former (two-qubit operations) are known to be noisier, it is in our interest to minimise their number. Theory of Makhlin transforms [13,14] shows us how to decompose any U_4 into local unitary gates and CNOT gates. It turns out that the unitary can be implemented (up to a phase) by using just two CNOT gates and six unitary gates. Translating the U_4 transformation into a circuit involves some highly nontrivial steps, which we have implemented using numerical methods (see the supplementary material). Our mathematica programs compute the required unitaries and produce a qasm program that can be implemented on the IBM machines. The quantum logic circuit implementing the entangling strategy is shown in figure 2.

3.2 Results

Let us summarise the results of the study. We have implemented our idea at three levels:

- (a) A mathematica program which numerically computes and compares the relative entropy for the

two computational strategies – the direct and the entangled.

- (b) A simulation on the `ibmqx4` computer which mimics the actual quantum experiment by incorporating some of the specific engineering details.
- (c) An actual quantum experiment on the `ibmqx4` computer.

Table 1 summarises our results. The relative entropies corresponding to the direct and entangling strategies, S_{reldir} and S_{relelt} , and their difference, $S_{\text{reldiff}} = S_{\text{relelt}} - S_{\text{reldir}}$, are displayed corresponding to the mathematica program (Theory), IBM simulator (Sim) and IBM experiment (Expt) in the table.

We notice that our mathematica programs are in close agreement with the results of the simulations on the `ibmqx4` quantum computer. However, there is a considerable discrepancy between the results of the real quantum `ibmqx4` experiments and the `ibmqx4` simulations or the results of the mathematica programs. This discrepancy could be due to limitations coming from gate infidelity and decoherence. Thus, the noise from these two sources masks the quantum advantage effect that we expect from our theoretical studies [6].

We graphically capture the main result of our study in figure 3 where we plot the $\text{Log}[\text{Likelihood}]$ vs. the

number of qubits which clearly demonstrates the superiority of the entangled strategy over the direct strategy. Thus, there is a clear quantum advantage in choosing the entangled strategy instead of the direct one.

3.3 Comparison with earlier work

There has been work in the area of state discrimination through ancilla measurement [15]. Samal *et al* [15] have focussed on the discrimination of Bell states which plays an important role in, for instance, teleportation and secret sharing. They have demonstrated this using an NMR-based experimental realisation of a protocol for Bell state discrimination. The state discrimination is performed through two independent measurements on one ancilla qubit which leaves the Bell states unchanged.

In the present paper, we consider two impure density matrices ρ_A and ρ_B and compare the efficacy of discriminating these two states via a direct strategy and an entangled strategy and compare our theoretical expectations against simulations and experiments on the `ibmqx4` computer.

4. Conclusion

In this paper, we described how likelihood theory works in the quantum regime and show how these ideas can be implemented on a quantum computer. The main message is that entanglement can be used as a resource to improve our power to discriminate between quantum states of qubits. Entanglement is crucially used here, because it is our ability to use nonseparable bases that gives us the quantum advantage.

The scheme we propose has been implemented on the `ibmqx4` computer. Shukla *et al* [16] have done a detailed quantum process tomography of each directly implementable quantum gate used in the IBM quantum processors to compute gate errors in the superconductivity-based quantum computers introduced by IBM and compared the quality of these gates with the corresponding gates implemented using other technologies. We have performed both simulations and real experiments on this computer. We notice that the simulation results are in agreement with our theoretical expectation of quantum advantage from entanglement in the context of quantum likelihood theory. However, the experimental results deviate from our theoretical expectation. This is very likely due to experimental noise, two possible sources being gate infidelity and decoherence. This noise masks the theoretically expected quantum entanglement advantage. This gap between the promise of quantum computers and the actual performance points

Table 1. Comparison of theory, simulation and experiment for S_{reldir} , S_{relelt} and S_{reldiff} .

	Theory	Sim	Expt
S_{reldir}	4.506	4.519	4.043
S_{relelt}	4.723	4.696	1.660
S_{reldiff}	0.217	0.177	-2.382

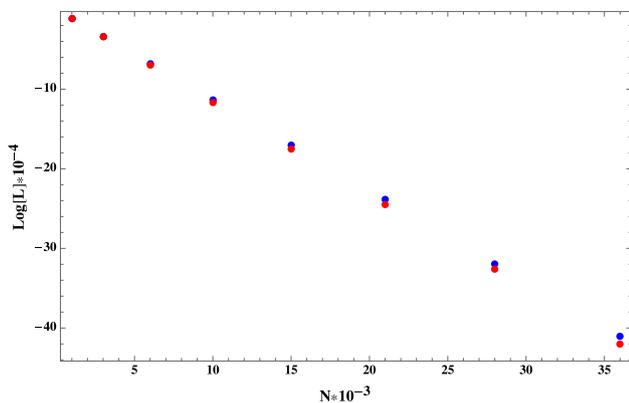


Figure 3. The log of the likelihood as a function of the number of runs. The red dots represent the result of using the entanglement strategy and the blue ones represent the result of using the direct strategy. The red dots show a steeper negative slope. This shows that the likelihood of the state being ρ_B decreases faster as the number of qubits measured increases.

to the need for improving the existing hardware. We expect that in future this gap may be closed by improving the design of quantum computers.

Acknowledgement

The authors acknowledge the use of the ibmqx4 quantum computer and the IBM Q16 computer in Melbourne.

References

- [1] S Amari, Information geometry and its applications, in: *Applied mathematical sciences* (Springer, Japan, 2016)
- [2] Y S Teo, *Introduction to quantum-state estimation* (World Scientific, Singapore, 2015)
- [3] A Chefles, *Contemp. Phys.* **41**(6), 401 (2010)
- [4] M Osaki, M Ban and O Hirota, *Phys. Rev. A* **54**, 1691 (1996)
- [5] S M Barnett and S Croke, *Adv. Opt. Photon.* **1**(2), 238 (2009)
- [6] K Shivam, A Reddy, J Samuel and S Sinha, *Int. J. Quant. Inf.* **16**, 1850032 (2018)
- [7] C W Helstrom, *Quantum detection and estimation theory* (Academic Press, Cambridge, 1976)
- [8] K M R Audenaert, J Calsamiglia, R Muñoz-Tapia, E Bagan, L Masanes, A Acin and F Verstraete, *Phys. Rev. Lett.* **98**, 160501 (2007)
- [9] C W Helstrom, *Phys. Papers A* **25**, 101 (1967)
- [10] S Abdel Khalek and S H A Halawani, *Pramana – J. Phys.* **85**, 1089 (2015)
- [11] J Shlens, [arXiv:1404.2000](https://arxiv.org/abs/1404.2000) (2014)
- [12] Supplementary, See Supplementary Material for programs and raw data
- [13] Y Makhlin, *Quant. Inf. Process.* **1**, 243 (2002)
- [14] V V Shende, I L Markov and S S Bullock, *Phys. Rev. A* **69**, 062321 (2004)
- [15] J R Samal, M Gupta, P K Panigrahi and Anil Kumar, *J. Phys. B* **43**, 095508 (2010)
- [16] A Shukla, M Sisodia and A Pathak, [arXiv:1805.07185](https://arxiv.org/abs/1805.07185) (2018)