Neural network oracle for non-locality testing of noisy Renou distribution

Bell's [1] initial study in 1964 has blossomed into a mature science. There were interests not only from a fundamental physics perspective, but also promised insight into the foundations of quantum theory and provide a supreme security in information processing tasks.

In 2012, Fritz [2] introduced distributions on networks that required no inputs for choice of measurement, yet analogous to the standard Bell scenarios. In-fact, he cleverly embedded the CHSH test into the triangle network, now known as the Fritz distribution.

However, this is not quite a satisfactory result. We know already that the independence of the sources causes the problem to lose its convexity. We expect to see different forms of non-locality arise out of this fact, that are completely distinct from the standard Bell scenario, as they would represent a different type of information resource. Secondly, the Fritz embedding can be done with just one entangled pair and a separable measurement basis. We wish to demonstrate examples with more entangled sources and measurements. To this end, Fritz issued the challenge of demonstrating genuine network non-locality, a notion motivated by the discussion above. One such example was proposed by Renou [3-5].

The Renou Distribution

In [3] Renou introduced a distribution on the triangle network that required the use of three entangled sources and partially entangled measurement basis, as a function of a parameter. Let's consider the token counting scenario below.

Consider the triangle scenario shown ahead.

Each source distributes the quantum state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$

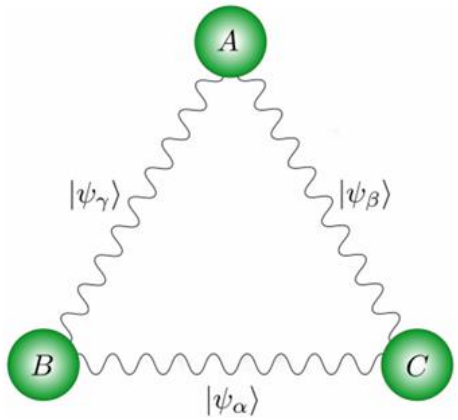

FIG. 1. Triangle network with three independent sources and measurement nodes

And each party performs measurement in the basis given by $\{\bar{0} = |00\rangle\langle 00|,\ \bar{1} = |01\rangle\langle 01| + |10\rangle\langle 10|, \bar{2} = |11\rangle\langle 11|\}$, where $\bar{n}$ counts the number of tokens received. Whenever the outcome $\bar{1}$ is obtained the party who obtained it performs a second, dichotomic measurement given by $\{|\overline{1_0}\rangle = u|01\rangle + v|10\rangle$, $\{|\overline{1_1}\rangle = -v|01\rangle + u|10\rangle\}$. We consider values of u^2 in the range of [0.5,1]. There is an alternative setup of a colour matching game that generates the same distribution, titled the RGB4 distribution. It was shown in [3], to be non-local for u^2 in $[0.785, 1]$ through a theoretical treatment of the problem using general properties of probability functions and the condition of rigidity, respected by token counting and colour matching games. It was later conjectured in [6], that the distribution is non-local for almost all values of u^2 except $\{0.5, 0.785, 1\}$. This was based on their numerical treatment of the problem. Soon after, in [7] the non-local range of the parameter was extended to include the set $[0.5631, 0.6563]$ through the clever use of an inflation argument. Furthermore, this was only a finite level of the inflation. It is then further conjectured that the real non-local range is even bigger. Given that not much of the range of the parameter is left to be local presently, it is probable that the entire parameter range apart from three points is non-local.

In real experiments, we expect noise to show up. The noise-resistance robustness of multiple distributions on the triangle network is are tested in [6]. They use two types of noise models in their testing, arising from the source and the measurement process.

Noise from the source is characterised using a visibility parameter v such that

$$\rho_{noise} = \left(v|\psi\rangle\langle\psi| + (1-v)\frac{1}{4}\right)^{\otimes 3}$$

The other model introduced a detector inefficiency given as a modification of the measurement basis into the following POVM

$$|\bar{n}\rangle\langle\bar{n}|_{noise} = v|\bar{n}\rangle\langle\bar{n}| + (1-v)\frac{1}{4}$$

It is observed in [6] that the two noise models, behave essentially the same. They stay local for small values of v till there is sudden transition and the distance increases all the way up till $v = 1$ (for a choice of measurement parameter in the non-local range). It is reported in [6] that the transition into the non-local regime occurs at $v^* \approx 0.9$ for both cases. Do note also the spiky nature of the graph in [6]. While it is an artifact of the code and also present in our runs too, we have obtained smoother curves that reveal that our initial estimate of the non-local limit for visibility is incorrect.

Methodology and Code

This study develops the ideas presented in [6]. Here, we will briefly describe the ideas of the first paper. The scenario is modelled using three feedforward neural networks. Each of these networks are given a random variable and asked to generate local response functions. We generate the combined probability distribution by averaging over many trails. The machine is trained to generate local responses such that the resultant combined distribution matches a target distribution. Because of the fact that classical random variables are being routed to the networks in an appropriate manner, which are asked to generate local response functions, the resultant combined distribution is guaranteed to be in the local set. Essentially the networks are asked to find a local decomposition for a given target distribution.

For this study, we have used the code from [6]. Noise analysis was one of the best utilities of the neural network. We were interested only in the Renou distribution. The numerical approach is especially effective as the proof of the non-locality of the distribution relied on symmetries that don't exist in the slightest bit of noise, thus being extremely challenging to investigate theoretically.

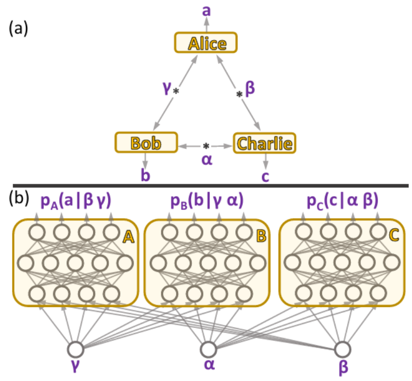

FIG. 2. Neural network structure

We implement a combined noise model for the Renou distribution and train it for various parameter values. We report improved results for previously simulated cases and the features of the new combined noise model.

Results

Firstly, we present the noise analysis results. Following are plots for the Euclidean distance D_e vs v_{global}, the source visibility.

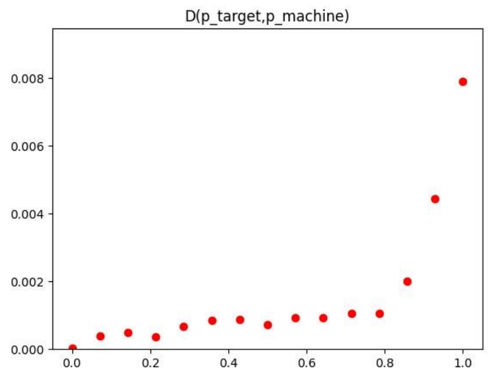

FIG. 3.a. $u^2 = 0.85$, $v_{\text{local}} = 1.00$

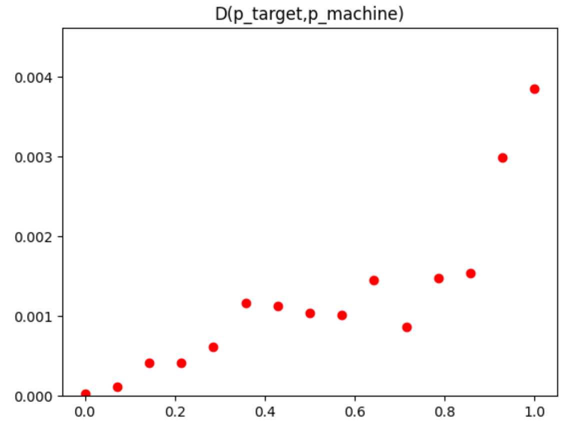

FIG. 3.b. $u^2 = 0.85$, $v_{\text{local}} = 0.95$

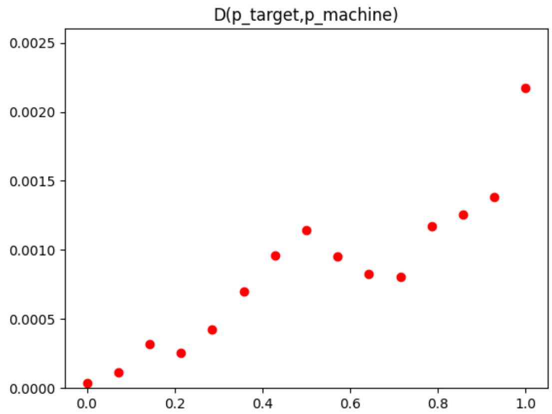

FIG 3.e. $u^2 = 0.85$, $v_{\text{local}} = 0.80$

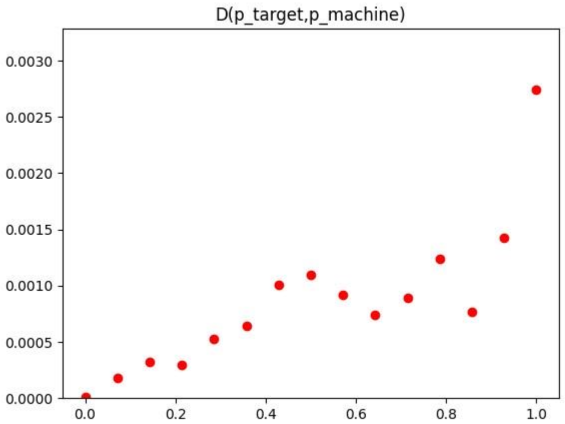

FIG 3.c. $u^2 = 0.85$, $v_{\text{local}} = 0.90$

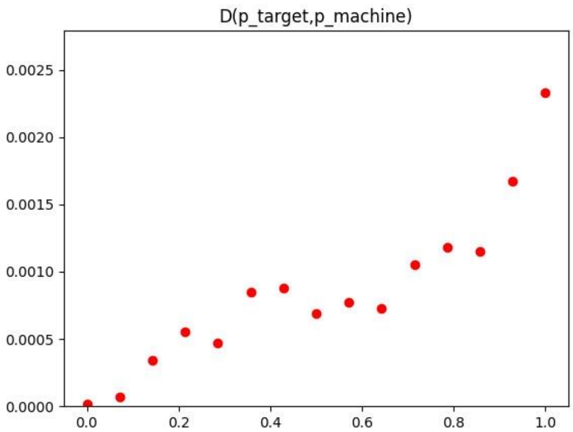

FIG. 3.d. $u^2 = 0.85$, $v_{\text{local}} = 0.85$

We also studied the variation of D_e vs u^2, tested for different values of v_{global} and v_{local}. Our results are in good agreement with [6]. Observe the dip near $u^2 = 0.785$ in all of the graphs below.

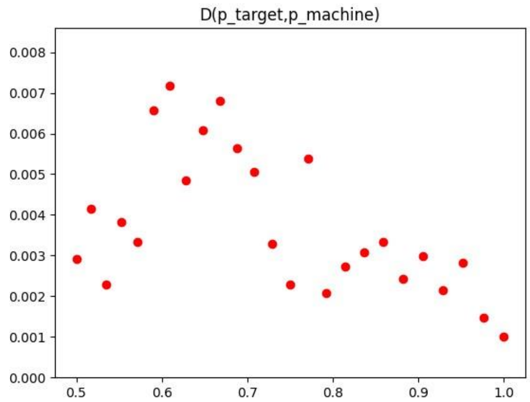

FIG. 4.a. $v_{\text{global}} = 0.95$, $v_{\text{local}} = 1.00$

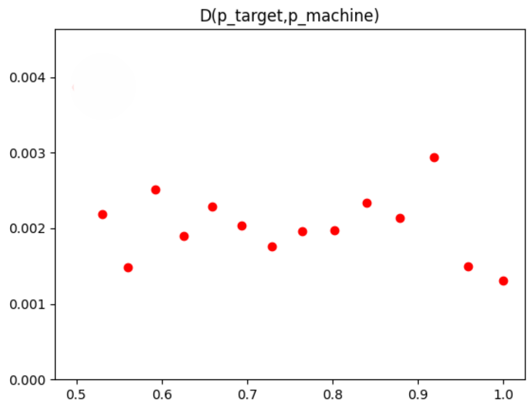

FIG. 4.b. $v_{\text{global}} = 1.00$, $v_{\text{local}} = 0.85$

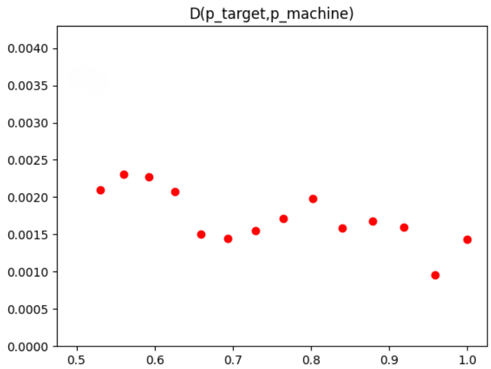

FIG. 4.c. $v_{\text{global}} = \sqrt{0.8}$, $v_{\text{local}} = \sqrt{0.8}$

Inferences

Surveying the data that has been generated we present a couple of inferences. Before we discuss our conclusions, we briefly describe some of the nuance of the analysis.

Because the machine is tasked with learning the best possible local strategy, it generates functions that minimizes the Euclidean distance between the target and the generated distribution defined as the following

$$D_e = \sum_{a,b,c} \left| p_t(a,b,c) - p_g(a,b,c) \right|^2$$

Theoretically, the machine, barring the possibility of getting stuck at a local minimum, should be able to learn the distributions that are truly local and satisfy $D_e = 0$, while the machine would be unable to learn a non-local distribution and D_e will take a finite value. However, in practise, the machine can only get "close" to the target distribution. D_e will be finite even for local distributions. Thus, to determine the non-local limit, we not only consider the value of D_e but also consider the point at which we observe the values of D_e quickly start increasing, i.e. the point at which the slope changes sharply. Thus, the non-local limit is set at that point at the distance at that point is the non-local distance limit.

Analysing the data above we make two conclusions. Firstly, observe the graph where local visibility parameter is set to unity. Observe how the slope of the graph sharply picks up near $v^* \approx 0.8$. We decide that the non-local limit is 0.8. Also observe that the distance at that point is around 0.002. Thus, the non-local limit is $D_e^* \approx 0.002$, where we recognise 0.002 as machine training error.

Secondly, looking at the results for different visibility value, we are motivated to define the quantity of a stacked visibility for the joint noise model simply as

$$v_{stacked} = v_{local} * v_{global} = v_{deteff} * v_{vis}$$

It is observed that the non-locality of a distribution is characterised by this property. If we assume $v_{local} = 1$, then from the discussion above, at non-local limit we would have $v_{stacked}^* = 0.8$ and $D_e^* = 0.002$.

Taking into account the results for all the other values of visibility, we conclude finally that for non-local distributions

$$v_{stacked}^* \approx 0.8 - 0.83$$

and,

$$D_e^* \approx 0.0015 - 0.002$$

Discussions and Suggestions

Our findings are encouraging for experimental realisation of network non-locality. Firstly, we propose a revised value for the non-local noise limit, allowing a much bigger noise allowance. Secondly, the observation of a stacked visibility is also favourable for experiment. This is because the stacked model implies essentially a linear addition of two types of noises.

The code has a significantly easier time learning the Fritz and the Elegant distribution models with small network sizes. This allows one to obtain reliable results on average computers within 10-15 minutes. The Renou distribution, however, requires a much larger network size to train which may take multiple days on average computers. Even using cloud resources, the run time still exceeds 3 hours regularly to produce good quality results. Apart from increasing the network size, increasing the variance of the input random variable set also aids in learning tougher distributions like the Renou distribution. For the Renou distribution, we took a network with depth and width as 10 and 60 respectively. We train each trail for 10000 epochs with a batch size of 8000, setting the variance of the data set to 0.35. More robust learning schemes and structures may have great value not only for the cases we have considered above but also for bigger more complicated network structures.

References

[1] J. S. Bell, *Physics Physique Fizika 1, 195 (1964)*.

[2] T. Fritz, *New J. Phys. 14, 103001 (2012).*

[3] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, *Phys. Rev. Lett. 123, 140401 (2019)*

[4] Marc-Olivier Renou and Salman Beigi, *Phys. Rev. Lett. 128, 060401 (2022)*

[5] M.-O. Renou and S. Beigi *(unpublished)*

[6] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, npj Quantum Inf. 6, 70 (2020)

[7] A. Pozas-Kerstjens, M.-O. Renou, and N. Gisin *(unpublished)*