National University of Singapore

Department of Physics

Possible Statistics from Bell Violations

Author:
Goh Koon Tong

Supervisor:
Professor Valerio Scarani
Mentor:
Le Phuc Thinh

Thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Science (Honours)

April 2014

Abstract

The Bell experiment [3] was designed to determine if local realistic models can account for certain experimental outcomes. The violation of Bell inequalities for a given set of experimental data would prove otherwise: the data cannot be pre-determined by a local realistic model [7], and in particular the outcome possess *private randomness*. The amount of private randomness present in the outcomes depends on the extent of Bell violation measured by the CHSH value. However, in order to obtain the CHSH value, one often depends on the assumption that the measured state is independently and identically distributed (IID). The relaxation of the IID assumption may pose problems especially in the case of finite-size statistics [2, 9, 15]. Indeed, there exist non-IID situations which hinders the certification of Bell violation. We will be studying those situations using a data analysis protocol which allows the relaxation of IID assumption, known as the prediction-based ratio (PBR) protocol [15]. By making the appropriate modifications to this protocol, it is possible to quantify the amount of private randomness present in the outcome of non-IID Bell experiments.

Acknowledgements

I would like to take this opportunity to express my heartfelt gratitude to my project supervisor, Professor Valerio Scarani, for his encouragement and guidance which kept me going and kept my project alive respectively. I would also like to thank him for the opportunity to work in a productive, fun and friendly environment such as connect.

Next, I would like show my appreciation to my mentor, Le Phuc Thinh, for his undying dedication and patience to get my concepts right and keep my work at least up to standard.

Without the assistance from Valerio and Thinh, the probability of this thesis being completed is comparable with that of a cat successfully tunneling through a wall.

Also, I would like to thank Jean-Daniel Bancal for the productive discussions which contribute significantly to my project.

Last but not least, I would like to thank my family and friends for putting up with my unexplained disappearance, mood swings and a sudden high demand for late night suppers and to be always there for me.

Contents

Chapter 1

Introduction

The essence of science lies within its reproducibility through observations and experimentations. Hence, it is crucial that the data extracted from scientific experiments are interpreted correctly. The study of statistics is the tool for scientists to do just that.

The probabilistic nature of the measurement outcomes from quantum system demands an infinite data sample from the system with independently and identically distributed (IID) state in order to determine the probabilities of occurrence of all possible outcomes. Only then, a complete picture of the studied system can be achievable.

In the field of statistics, the state parameter estimations from finite data sampled from IID processes are well studied and known. However, if there is a drift in the state of the measured system or a change in the environment of the experiment, the IID assumption will fail to hold. The task of studying these systems were made to be more challenging by the fact that sampling infinite data from such system could be infeasible or physically impossible.

These considerations have a direct impact on the field of quantum information. Thus, there have been many recent theoretical works done on the topics like quantum state estimation [12], determination of non-classical correlations with information-theoretic distance [10] and certification of Bell violation [2, 8, 9, 15] which considers cases of finite statistics.

The discussions of this paper will be focused on the Bell experiment [3] with relaxation on the IID assumption, due to its many applications. The proposition of Bell experiment has its historical purpose in falsifying local realism [7]. In recent years, the applications of Bell experiment extend to

the generation of private randomness [11] and quantum cryptography [1].

In following chapters, discussions of Bell experiment, its application as random number generator and the certification of Bell violation and private randomness generation in non-IID Bell experiment will be presented.

Firstly, the notion of randomness and the quantity that measures randomness, min-entropy, will be introduced. Then, a distinction will be made between random and pseudo-random processes. Subsequently, the idea of local realism will be brought forth and the construction of Bell experiment to falsify local realistic models will be mentioned. Next, it will be shown that indeed private randomness can be extracted from Bell violation.

The framework of hypothesis testing will be briefly discussed together with the data analysis protocols for Bell experiment which were introduced by Zhang and his coworkers [15]. Finally, the PBR protocol will be applied to certain non-IID situations and modified to quantify private randomness from non-IID Bell experiment.

Chapter 2

Preliminaries

In this chapter, all relevant background knowledge required to appreciate the work will be introduced.

In the first section, the notion of randomness will be introduced with classical example and quantified using min-entropy. Then, we will show that randomness is not an absolute quantity as it depends on the amount of relevant information which is available to the observer. Finally, the distinction between randomness and pseudo-randomness will be demonstrated.

In the second section, the notion of local realism and its role in physics will be formally introduced. Next, we will demonstrate how Bell experiment is able to put local realistic model to the test. In order to achieve that, it is necessary to determine the predictions of the experimental outcomes drawn by local realistic models and quantum theory. By making comparison of different predictions, we will be able to construct a criterion, which is known as Bell inequalities, whereby local realistic models can be falsified.

2.1 Randomness

2.1.1 Introduction to Randomness

Consider the scenario of rolling a six-sided die, we define X as the outcome of the die roll. The possible outcomes of X is 1,2,3,4,5 and 6, denoted by $X \in \{1, 2, 3, 4, 5, 6\}$. Then, we assume that the probabilities of any outcomes occurring to be non-zero. Here, X is a random process because it is impossible to predict with definite confidence the outcome of each trial of X.

By rolling the die repeatingly indefinitely, it is possible to obtain the probability distribution, ρ, of the outcomes. The probability distribution, ρ, can be represented by a vector as shown below:

$$\rho(X) = (\Pr(X = 1), \Pr(X = 2), \cdots, \Pr(X = 6)) \tag{2.1}$$

Here we denote $\Pr(X = x)$ as the probability that the event, X, results in the outcome, x.

The notion of randomness describes the unpredictability of an event. A random process is required to have more than one possible outcomes and the probability for each outcome to occur has to be non-zero. In this way, it is impossible to make prediction on the outcome with definite confidence.

2.1.2 Quantifying Randomness: min-entropy

There exist many quantities in the literature which quantifies randomness and the commonly used quantity for Bell experiment is the min-entropy, $H_{\min}$. In this section, we will introduce the min-entropy and demonstrate how min-entropy is a good quantity which describes randomness.

Consider a guessing game where Alice will roll a six-sided die and Bob has to guess the outcome, X. Given that Bob has knows the probability distribution of the outcome, $\rho(X)$, what will be Bob's strategy to maximise his chance of winning this game? Obviously, the best strategy that Bob could adopt is to choose the most probable outcome.

In this case, the probability that Bob wins the guessing game is the probability that the most probable outcome occurs. This probability is known as the guessing probability, denoted by $\Pr_{guess}(X)$.

It is apparent that a minimal $\Pr_{guess}$ will describe an event with maximum randomness. On the other hand, a maximal $\Pr_{guess}$, which has a value of 1, describes a deterministic event. A good measure of randomness should therefore taken into account of the $\Pr_{guess}$ of the event.

Next, if two independent trials of X were to be conducted, the $\Pr_{guess}$ of the combined event, X_1, X_2, is given by:

$$\Pr_{guess}(X_1, X_2) = \Pr_{guess}(X_1)\Pr_{guess}(X_2) \tag{2.2}$$

However, since randomness describes the uncertainty of an event, a good

measure of randomness should have an additive properties between 2 independent events.

In order to fulfil these properties of the randomness of events, we define the min-entropy, $H_{\min}$, of an event, X, as follows:

$$H_{\min}(X) = -\log_2[\Pr_{guess}(X)] \tag{2.3}$$

The unit of min-entropy is bit. The logarithmic base of 2 is chosen because in information theory, a bit of information has 2 possible states, namely 0 and 1. Therefore, given a maximally random bit, the $\Pr_{guess}$ will be given by $\frac{1}{2}$ which translate to a min-entropy, $H_{\min}$, of 1 bit.

2.1.3 Randomness for whom?

Randomness is relative to the different amount of relevant information available to the observer and this can be demonstrated by the following example. Now, consider that Alice hid a ball underneath 1 of the 3 inverted cups, and Bob has to guess under which cup is the ball hidden. Assume that the cups are perfectly opaque and there is no way that Bob has the information of the whereabouts of the ball, other than it being in 1 of the 3 cups . Now, we denote these information available to Bob to be B.

Since Bob's best available strategy is to make a wild guess, which implies that guessing probability of outcome, X, conditioned on the information available to Bob is given by:

$$\Pr_{guess}(X|B) = \frac{1}{3} \tag{2.4}$$

Now, we define the conditioned min-entropy of the event, X, given by the information B, to be as follows:

$$H_{\min}(X|B) = -\log_2[\Pr_{guess}(X|B)] \tag{2.5}$$

$$= -\log_2\left[\frac{1}{3}\right] \approx 1.58 \text{ bits} \tag{2.6}$$

However, since Alice hid the ball, she has complete information about the ball's whereabouts. Hence, if we denote the information available to

Alice as A, the min-entropy conditioned on A is given by:

$$H_{\min}(X|A) = -\log_2[\Pr_{guess}(X|A)] \tag{2.7}$$
$$= -\log_2[1] = 0 \text{ bit} \tag{2.8}$$

From the above example, it is clear that the same event viewed from different point of views give rise to different results of conditioned min-entropy. Hence, it is conclusive that randomness does depend on the amount of information available.

These events which appear random due to the lack of information are known as pseudo-random events. With sufficient information, all pseudo-random events are deterministic. If we denote any pseudo-random events as X_p and all possible information obtainable from the system by C, we can write:

$$H_{\min}(X_p|C) = 0 \tag{2.9}$$

For the purpose of this paper, the generation of pseudo-random outcomes is not of any interest. The discussion in the following chapters will be focused on the random outcomes generated by Bell experiment, X_B, whereby the following condition is met.

$$H_{\min}(X_B|C) > 0 \tag{2.10}$$

2.2 Bell experiments

2.2.1 Local Realism

Local realism [7] is one of the many possible physical models which seeks to explain the behaviour of nature. The notion of local realism consists of two components, namely locality and realism.

- Locality is a physical concept which states that the rate at which information propagates over space is upper bounded by the speed of light. According to locality, there are no instantaneous cause and effect over any non-zero spatial distance as information takes time to travel over space.
- Realism requires the properties of all physical objects to be real regardless of the measurements made on these objects. In other words,

the act of measurement merely uncover the pre-established value of the measured property. By taking a measurement on a physical object, the state of the object is not altered in anyway.

The local realistic model works flawlessly in the realm of classical physics. In classical physics, any faster-than-light speed propagation is strictly prohibited by special theory of relativity which obeys the locality constraint. Additionally, in classical physics, all degrees of freedom of a physical object is governed by its equation of motion. Therefore, with the knowledge of the appropriate equations of motion, one will be able to make prediction of the outcome of any measurement made on the physical object. This statement could only be true if the concept of realism truly describes nature. Thus, in the classical picture, local realism is a valid model which describes nature.

However, as it will be evident later, the local realistic model is not compatible with the quantum theory and there exist experiments, which are known as Bell experiments, which seek to determine if nature can indeed be described by the local realistic model.

2.2.2 Bell Experiment

In 1964, John Stewart Bell came up with an experimental scheme [3] which can potentially put the validity of the local realistic model to the test. Subsequently, it is experimentally implemented in 1969 by John Clauser, Michael Horne, Abner Shimony and Richard Holt (CHSH) [6].

The experimental scheme adopted by CHSH involves a bi-partite system, whereby both parties, typically given the name Alice and Bob, are space-like separated. Then, 2 particles produced by a source will then propagate to each party. Both Alice and Bob are required to make a measurement on their given particle. Each party is allowed to choose 1 out of the 2 different, possible types of measurements and each measurement will give 1 out of the 2 different, possible outcomes. This particular scenario of Bell experiment with 2 parties, 2 measurement settings and 2 outcomes is also known as the CHSH experiment.

Here, we denote the measurements of Alice and Bob to be i and j respectively and we denote the 2 possible measurements to be 1 and 2, as such $i, j \in \{1, 2\}$. Similarly, we denote the outcomes of the measurements made by Alice and Bob to be a and b respectively and we denote the 2 possible

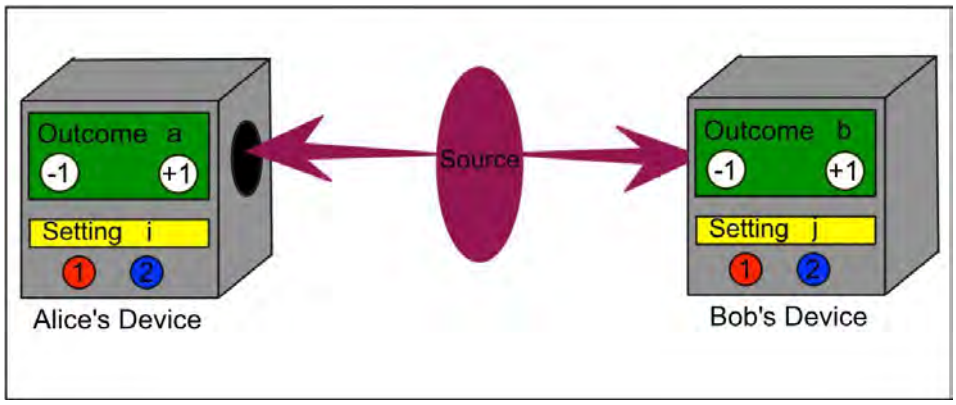

Figure 2.1: The diagram of a CHSH experiment. Each particle from the source will enter the devices of Alice and Bob. Then, Alice and Bob will select a setting (either 1 or 2) for the measurement made on the particle that entered the device. Finally, the outcome (either -1 or 1) will be displayed on the device.

outcomes to be -1 and 1, as such $a, b \in \{-1, 1\}$.

Now, we will introduce a local variable, λ, which provides strategies for the measured particles to behave accordingly. One can see λ as a set of strategies whereby when one inputs the values of i and j, λ will produce the outcomes, a and b. In other words, the knowledge of λ allows one to predict the outcome. This can be illustrated by the following equation:

$$\Pr_{\text{guess}}(\ a\ b \mid \lambda\ ) \overset{\text{LR}}{=} 1 \tag{2.11}$$

That being said, the direct result of a Bell experiment is not the guessing probability but the conditional probabilities, $\Pr(\ a\ b \mid i\ j\ )$, which can be written as:

$$\Pr(\ a\ b \mid i\ j\ ) = \int d\lambda\ \rho(\ \lambda \mid i\ j\ ) \Pr(\ a\ b \mid i\ j\ \lambda\ ) \tag{2.12}$$

where $\rho(\ \lambda \mid i\ j\ )$ is the probability distribution of λ given a specific values of i and j. The above expression is nothing other than the Bayes' theorem written for continuous variable, λ, with an integral instead of the convention summation for discrete random variables.

Since both parties are space-like separated, by invoking the locality constraint, we can infer that no information can travel between Alice and Bob. Taking this into consideration, we can write down 3 constraints on the system, namely the outcome independence, measurement independence and the no-signalling constraint.

The outcome independence comes about because each party will only have information based on the particle he/she received from the source and the measurements made on the particle. Hence, we can write:

$$\Pr(\ a\ b \mid i\ j\ \lambda\ ) = \Pr(\ a \mid i\ j\ \lambda\ )\Pr(\ b \mid i\ j\ \lambda\ ) \tag{2.13}$$

Likewise, using the same argument, the information of one party's measurement on his/her particle does not travel to the other party. Therefore, the no-signalling constraint can be mathematically written as:

$$\Pr(\ a \mid i\ j\ \lambda\ ) = \Pr(\ a \mid i\ \lambda\ ) \tag{2.14}$$

$$\Pr(\ b \mid i\ j\ \lambda\ ) = \Pr(\ b \mid j\ \lambda\ ) \tag{2.15}$$

Finally, since the measurement is made after the particles are produced, λ has to be measurement independence. This can be expressed by:

$$\rho(\ \lambda \mid i\ j\ ) = \rho(\ \lambda\ ) \tag{2.16}$$

Now, using the 3 constraints that we have introduced earlier, we can rewrite the equation for $\Pr(\ a\ b \mid i\ j\ )$ to:

$$\Pr(\ a\ b \mid i\ j\ ) = \int d\lambda\ \rho(\ \lambda\ )\Pr(\ a \mid i\ \lambda\ )\Pr(\ b \mid j\ \lambda\ ) \tag{2.17}$$

Under the assumptions of local realism, the outcomes of the measurements are required to have definite values, able to be determined via some form of equation of motion, prior to measurements. In other words, the particles measured by Alice and Bob are required to have a pre-established agreement on the outcomes of the measurements. This deterministic nature of the particles meant that with the knowledge of λ will give us a definite value of a and b for its corresponding measurements i and j. Therefore, the probabilities $\Pr(\ a \mid i\ \lambda\ )$ and $\Pr(\ b \mid j\ \lambda\ )$ can only take up the values 0 or 1 depending on the measurements i, j and variable λ. Mathematically, we can write:

$$\Pr(\ a \mid i\ \lambda\ ) \overset{\mathsf{LR}}{=} \delta(a - f(i,\lambda)) \tag{2.18}$$

$$\Pr(\ b \mid j\ \lambda\ ) \overset{\mathsf{LR}}{=} \delta(b - g(j,\lambda)) \tag{2.19}$$

Bell Inequalities

From the CHSH experiment, it is desirable to obtain an expression from the measurement outcomes which values can be bounded by the local realism assumption. In this way, if a violation of such bound can be observed experimentally, one can reject local realism. This bound is known as the Bell inequality.

The CHSH measurement expression is a linear combination of the correlation functions, $E_{i,j}$ of the outcomes a, b given the measurement settings i, j. The correlation function can be written as:

$$E_{i,j} = \Pr(a = b|i,j) - \Pr(a \neq b|i,j) \tag{2.20}$$

Recall that the measurement outcomes $a, b \in \{-1, 1\}$, which implies that the product of the outcome, ab, is related to its correlation in following expression:

$$ab = \begin{cases} 1 & \text{if } a = b \\ -1 & \text{if } a \neq b \end{cases}$$

Hence, the correlation function, $E_{i,j}$, can be rewritten to be:

$$E_{i,j} = \Pr(ab = 1|i,j) - \Pr(ab = -1|i,j) \tag{2.21}$$

$$= \langle (ab)_{i,j} \rangle \tag{2.22}$$

$$= \langle a_i b_j \rangle \tag{2.23}$$

$$\overset{\mathsf{LR}}{=} \langle a_i \rangle \langle b_j \rangle \tag{2.24}$$

In the third step of the derivation, the no-signalling constraint was invoked which requires the independence between a and j, similarly also between b and i. Also, in the last step of derivation, the local realism assumption requires the independence between measurement outcomes a and b as their only dependence are the measurement settings and strategy, λ.

After going through the required knowledge about the correlation func-

tion, the CHSH measurement expression can be defined as follows:

$$\text{CHSH} = E_{11} + E_{21} + E_{12} - E_{22} \tag{2.25}$$

Now, by applying the local realism constraint on the CHSH measurement, the current interest is to determine the upper bound of CHSH values. This can be shown in the following derivations:

$$\begin{aligned}
\max_{LR} \text{CHSH} &= \max_{LR}(E_{11} + E_{21} + E_{12} - E_{22}) && (2.26)\\
&\overset{\text{LR}}{=} \max_{LR}(\langle a_1\rangle\langle b_1\rangle + \langle a_2\rangle\langle b_1\rangle + \langle a_1\rangle\langle b_2\rangle - \langle a_2\rangle\langle b_2\rangle) && (2.27)\\
&= \max_{LR}(\langle a_1\rangle[\langle b_1\rangle + \langle b_2\rangle] + \langle a_2\rangle[\langle b_1\rangle - \langle b_2\rangle]) && (2.28)\\
&= 2 && (2.29)
\end{aligned}$$

By the same argument, the minimum CHSH value is determined to be -2. Hence, the range of values of CHSH as predicted by any local realistic model is given by:

$$-2 \leq \text{CHSH} \leq 2 \tag{2.30}$$

However, without losing any generality, it is useful to just consider the positive values of CHSH. By doing so, we will arrive at the famous CHSH inequality, which is given by:

$$E_{11} + E_{21} + E_{12} - E_{22} \leq 2 \tag{2.31}$$

The above criterion has to be met by any local realistic model. Of course, there exist other CHSH values which falls outside the range of the inequality. It is apparent that the range of CHSH values without any constraint is given by $-4 \leq \text{CHSH} \leq 4$. In the next section, the violation of the Bell inequality will be demonstrated via quantum theory.

Violating Bell Inequalities

In quantum theory, we can write the CHSH measurement expression as an operator denoted by $\hat{S}$ and is given by:

$$\hat{S} = \hat{A}_1 \otimes \hat{B}_1 + \hat{A}_1 \otimes \hat{B}_2 + \hat{A}_2 \otimes \hat{B}_1 - \hat{A}_2 \otimes \hat{B}_2 \tag{2.32}$$

In this case, the CHSH inequality will be given by:

$$\left\langle \hat{S} \right\rangle \leq 2 \tag{2.33}$$

where $\left\langle \hat{S} \right\rangle$ =Tr($\hat{\rho}\hat{S}$) with $\hat{\rho}$ representing the density matrix of the system of the 2 particles measured by Alice and Bob and Tr is the trace operation. Since local realistic model requires the upper bound of $\left\langle \hat{S} \right\rangle$ to be 2, we will be able to reject local realism by obtaining experimental data that proves the relation $\left\langle \hat{S} \right\rangle > 2$.

Now, we will like to find out whether there exist values of $\left\langle \hat{S} \right\rangle$ predicted by quantum theory which violates the CHSH inequality. The obvious approach is to determine the upper bound value of $\left\langle \hat{S} \right\rangle$ as predicted by quantum theory. It is known that the constraint of $\left\langle \hat{S} \right\rangle$ by quantum theory is given by the Tsirelson bound [5].

To derive the mathematical expression of Tsirelson bound, we will now take the square of the operator $\hat{S}$. Now, recall that the possible eigenvalues of $\hat{A}_i$ and $\hat{B}_j$ are ± 1. Therefore, the values of $\left\langle \hat{A}_i^2 \right\rangle$ and $\left\langle \hat{B}_j^2 \right\rangle$ have to be 1. Hence, we can conclude that A_i^2 and $\hat{B}_j^2$ are $\mathbb{1}$. Now, $\hat{S}^2$ can be expanded and simplified to the following expression:

$$\hat{S}^2 = 4\mathbb{1} \otimes \mathbb{1} - [\hat{A}_1, \hat{A}_2] \otimes [\hat{B}_1, \hat{B}_2] \tag{2.34}$$

Since we are only interested in determining the upper bound of the eigenvalue of $\hat{S}^2$, there is no need to evaluate the above expression, consideration of its maximum magnitude of its eigenvalue is sufficient. Let us consider the maximum magnitude of the eigenvalues of the commutator $[\hat{A}_1, \hat{A}_2]$.

$$\begin{aligned} \max|[\hat{A}_1, \hat{A}_2]| &= \max|\hat{A}_1\hat{A}_2 - \hat{A}_2\hat{A}_1| && (2.35)\\ &\leq \max|\hat{A}_1\hat{A}_2| + \max|\hat{A}_2\hat{A}_1| && (2.36)\\ &\leq 2\ \max|\hat{A}_1|\ \max|\hat{A}_2| && (2.37)\\ &= 2 && (2.38) \end{aligned}$$

In the second step of the derivation, recognise that the maximum eigenvalue of the difference between 2 operators corresponds to the scenario when the eigenvalues of the 2 operators have the opposite sign. Moreover, the mag-

nitude of the eigenvalue of the difference between 2 operators will be simply the sum of the magnitudes of the constituent eigenvalues. Also, in the third step of the derivation, we invoked the Cauchy-Schwarz inequality.

By similar arguments as the derivation above, the maximum magnitude of the eigenvalues of $[\hat{B}_1, \hat{B}_2]$ is also 2. Now, it is obvious that the maximum magnitude of eigenvalues of $\hat{S}^2$ is 8. Finally, by taking square root of this relationship, we will arrive at the famous expression of the Tsirelson bound:

$$\left\langle \hat{S} \right\rangle \leq 2\sqrt{2} \tag{2.39}$$

This inequality shows the possible values of $\left\langle \hat{S} \right\rangle$ according to quantum theory. It is apparent that according to quantum theory, there exist a range of values of $\left\langle \hat{S} \right\rangle$ which is forbidden by local realistic models. This implies that if the condition $2 < \left\langle \hat{S} \right\rangle \leq 2\sqrt{2}$ is experimentally observed, there is a violation of local realism.

The LR Polytope

Since the measurement settings values of i, j are chosen arbitrarily, the Bell inequalities under measurement settings permutations are equally valid. Thus, for a given Bell experiment there are multiple Bell inequalities bounding the set of probability distribution of settings and outcomes which can be accounted for by local realistic model. Amidst this mess, by interpreting this set of probability distribution geometrically, it clears up the mess creating by considering multiple equations without any intuition of the scenario.

Since there are 16 different settings-outcomes combinations, a settings-outcomes probability distribution can be interpret as a point residing in the real space, $\mathbb{R}^{16}$. Each coordinate of its position in $\mathbb{R}^{16}$ are in fact the probability of occurrence corresponding to the axis. Hence, there are 2 constraints on all outcome-settings probability distributions, x, that dictates the allowed subspace for the distributions to reside in.

1. Positivity of probabilities: $x_i > 0,\ i \in [1, 16]$
2. Convexity of probabilities: $\sum_{i=1}^{16} x_i = 1$

Geometrically, the above mentioned subspace is a "quadrant" of a 16-dimensional hyper sphere. In this subspace, there exists a set of probabilities

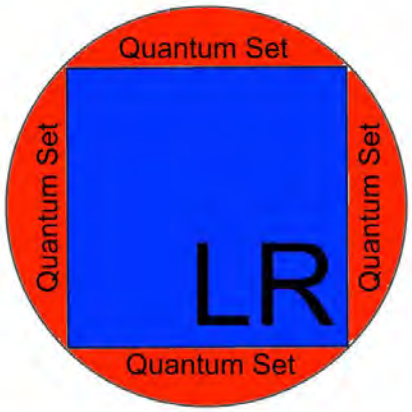

Figure 2.2: LR Polytope and Quantum Set in geometrical representation. The blue square denoted "LR" represents the LR polytope which contains all probability distributions which fulfils the local realism constraints. The red circle represents the quantum set which contains all probability distributions which can be accounted by quantum theory.

distributions constraint by local realistic models, denoted by LR. This LR set is defined by its linear constraints known as the Bell inequalities. In this geometric picture, a total of 8 Bell inequalities have to be considered. These Bell inequalities are as shown below:

1. $E_{11} + E_{21} + E_{12} - E_{22} \leq 2$
2. $E_{11} + E_{21} + E_{12} - E_{22} \geq -2$
3. $E_{11} + E_{21} - E_{12} + E_{22} \leq 2$
4. $E_{11} + E_{21} - E_{12} + E_{22} \geq -2$
5. $E_{11} - E_{21} + E_{12} + E_{22} \leq 2$
6. $E_{11} - E_{21} + E_{12} + E_{22} \geq -2$
7. $-E_{11} + E_{21} + E_{12} + E_{22} \leq 2$
8. $-E_{11} + E_{21} + E_{12} + E_{22} \geq -2$

These 8 linear constraints will form the borders of a polytope and this polytope contains all probability distributions which can be explained by local realistic models. For convenience, this polytope is defined to be the local realism (LR) polytope.

Additionally, recall that all settings-outcomes distribution that can be accounted for by quantum theory is defined by the Tsirelson's bound. Since the Tsirelson's bound has less constrain on the CHSH value, intuitively, the LR polytope will be included in the quantum set.

Figure 2.2 is a diagram which attempts to give an intuition of the geometric interpretation of Bell violation. This diagram is not a stereographic projection from $\mathbb{R}^{16}$ to $\mathbb{R}^2$, the square and circle are purely symbolic to aid the visualisation of hyper-dimensional objects. From the diagram, any settings-outcomes distribution that is found to be lying outside the LR polytope is said to violate Bell inequality.

Of course, the distribution is expected to lie within the quantum set. If it happens that the distribution is found to be lying outside the quantum set, quantum theory will not be able to explain such experimental outcome correlations.

In later chapters, this geometric interpretation of Bell violation will be used to explain data analysis protocol with much clarity than equations.

2.2.3 Randomness in Bell Experiment

One of the important applications of Bell experiment is the generation of random numbers. As shown in the previous section, by conducting the Bell experiment, it is possible to determine if the particles measured by Alice and Bob behave according to any local realistic model or not.

Since the assumption of realism demands a pre-determined value for any degrees of freedom prior to measurement, having the knowledge of that value would renders that quantity to be deterministic. Since all classical phenomenon can be explained by some local realistic models, it implies that all classical phenomenon are deterministic or, at best, pseudo-random.

Fortunately, according to quantum theory, there exist some states that produce correlations which violate the Bell inequality. In such scenarios, the local realism assumption breaks down which implies that the predetermination of these measurement outcomes are not allowed.

Since the measurement outcomes are only determined upon the moment of measurement, it is impossible to make a prediction of the outcomes with complete confidence. In this way, the Bell experiment is able to generate private random numbers. Hence, violating the Bell inequalities certifies that the outcomes produced are indeed private and random.

In order to relate the amount of randomness produced in a Bell experiment and Bell violation, consider an arbitrary pure state given by $|\psi>= \cos\theta|00> + \sin\theta|11>$ undergoing measurements of a Bell experiment. The maximum achievable CHSH value for such quantum state over all measurements can be found to be:

$$\text{maxCHSH} = 2\sqrt{1+\sin^2 2\theta} \tag{2.40}$$

In this problem, it is only required to consider the guessing probability of one party and in this example we choose to consider Alice's measurement. Recognise that the guessing probability of Alice's measurement is given by the probability of the most probable outcome over all possible measurements made by Alice. Given the state $|\psi>$, the most biased marginal probability is given by:

$$P(+|\hat{z}) = \frac{1}{2}(1+\cos 2\theta) \tag{2.41}$$

By solving the above two equations simultaneously, we will arrive at the guessing probability of the outcome from Alice's measurement, which is given by:

$$\Pr_{\text{guess}}(a|i) = \frac{1}{2}\left[1+\sqrt{2-\left(\frac{\text{CHSH}}{2}\right)^2}\right] \tag{2.42}$$

Therefore, the min-entropy of the outcome, a, measured by Alice can be computed and is given by:

$$H_{\min}(a|i) = \log_2\left(\frac{1}{2}\left[1+\sqrt{2-\left(\frac{\text{CHSH}}{2}\right)^2}\right]\right) \tag{2.43}$$

$$= 1-\log_2\left[1+\sqrt{2-\left(\frac{\text{CHSH}}{2}\right)^2}\right] \tag{2.44}$$

For any outcomes that do not violate Bell inequality (ie. CHSH = 2), it can be shown that $H_{\min}(a|i) = 0$. On the other hand, for any outcomes which violates the Bell inequality, the $H_{\min}(a|i)$ is always positive which implies that the outcomes are indeed private and random if Bell inequality is violated. On top of that, the amount of randomness obtained is related to the extent of Bell violation (ie. the CHSH value).

Chapter 3

Hypothesis Testing for Local Realism

Instead of writing Bell's inequality in the form of an averaged quantity over measurements and outcomes, it is possible to phrase Bell's inequality in the form of a hypothesis test and quantifying the Bell violation with p-values. This will allow the data analysts to relax the IID assumption of Bell experiment.

In this chapter, the framework of hypothesis testing will be briefly introduced. Also, several data analysis protocols for Bell experiment will be discussed. The content of this chapter is based on the main reference on which we build [15].

3.1 Hypothesis Testing

In many experiments in the laboratories, the common main objective is to prove or disprove a hypothesis using evidences supported by the experimental results. The extraction of experimental data from the setup can be seen as a form of statistical sampling from the population of interest. The data analysis done after every experiment aims to provide a good estimation of the population's parameters.

In the field of statistics, there exist a well structured procedure to prove or disprove hypothesis, which is known as hypothesis testing. In this section, the framework of the hypothesis testing will be introduced and briefly discussed.

The first step of hypothesis testing is to define the hypotheses. In this step, one has to define the null hypothesis, H_0 and the alternative hypothesis, H_a. It is common practice to construct the H_0 to be the hypothesis that the experimenter would like to reject. Thus, H_0 is often called the hypothesis of "no effect" by statisticians. On the other hand, the role of H_a is to provide an alternative explanation of the experimental results in the case when H_0 is rejected.

In the following statistical tests against local realism, the H_0 and H_a are given as follows:

H_0 : The results can be explained by local realistic model

H_a : The results cannot be explained by local realistic model

The next step of hypothesis testing is to select the test statistics. The test statistics is the distribution that the variable of interest takes up. Assuming each experimental trials are independently and identically distributed (IID) and the sample is randomly chosen and has a large sample size, then the normal distribution will be suitable to be the test statistics.

Once the test statistics has been obtained, the p-value can be computed. Assuming that the H_0 is true, the p-value is defined as the probability of obtaining the observed or more extreme statistics. In this paper, the protocols consider only the one-tailed test of the test statistics T. Thus, we will define p-value of T given the experimental data x to be the largest possible $\Pr(T \geq T(x))$ over all T constrained by local realism.

The final step is to decide whether or not to reject H_0. If the p-value is small, then the result is statistically significant and therefore we reject H_0. However, the distinction between large and small p-values were poorly defined. Hence, statisticians introduced the term "significance level", usually denoted by α, to define the boundary between "large" p-values and "small" p-values.

In hypothesis testing, when p-value is smaller than α, we reject H_0. The choice for the value of α is completely arbitrary. Historically, Sir Ronald Fisher set α to be 0.05 and this remains the convention in many existing literature.

Today, many statisticians felt that it is not necessary to fix a value for α. For a given p-value, the rejection of H_0 will be valid for all values of α such

that p-value $\leq \alpha$. Therefore, the p-value itself represents the lower bound of α such that the rejection of H_0 holds. Therefore, the results of hypothesis testing is presented with the p-value.

3.2 Conventional Method: SD-based Protocol

In this section, a data analysis methodology known as the standard deviation (SD)-based protocol will be introduced and the framework of this protocol which leads to the p-value will be discussed.

3.2.1 Rewriting the Bell inequalities

As the conventional expression of CHSH inequality contains expectation values of certain measurements and it offers very little intuition on the way it is supposed to be computed. Now, a new form of expression of Bell inequalities will be introduced to cut down on the ambiguity. Typically, a Bell inequality can be written as:

$$\langle S(x) \rangle \leq B \tag{3.1}$$

where $S(x)$ is the measurement function which takes up real values, x is the setting and outcome combination of a single trial measurement and B is the upper bound value of $\langle S(x) \rangle$ which can be accounted for by local realism.

For the case of the CHSH inequality, the measurement function, $S(x)$, will given by the mathematical expression:

$$S(x) = (1 - 2\delta_{i,2}\delta_{j,2})ab/p_{i,j}, \text{and } B = 2 \tag{3.2}$$

From the two above-mentioned equations, the familiar CHSH inequality can be obtained and is as shown below:

$$\langle S(x) \rangle = \sum_{i,j,a,b} p_{i,j,a,b}(1 - 2\delta_{i,2}\delta_{j,2})ab/p_{i,j} \tag{3.3}$$

$$= \sum_{i,j,a,b} p_{a,b}(1 - 2\delta_{i,2}\delta_{j,2})ab \tag{3.4}$$

$$= \langle a_1 b_1 \rangle + \langle a_1 b_2 \rangle + \langle a_2 b_1 \rangle - \langle a_2 b_2 \rangle \tag{3.5}$$

$$\leq 2 \tag{3.6}$$

Notice that $S(x)$, a function of experimental data x, is the test statistics in the hypothesis testing for local realism. Ideally, one can perform the experiment to obtain a value of $\langle S(x)\rangle$ to show that the value is greater than B, hence, violating local realism. However, in practice, we will expect our experimental results to deviate from the "true value" of the measurement.

Assuming that these trials of measurements are IID, a good experimental practice is to take a large number of readings. From the experiment, we obtain $x_1, \cdots, x_n$ from n trials. The conventional practice to estimate the value of $\langle S(x)\rangle$ is to take the average value of $S(x)$ of the individual trials over all n trials and is given by:

$$\bar{S} = \frac{1}{n}\sum_{k=1}^{n} S(x_k) \tag{3.7}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\sum_{i,j} p_{i,j}(1 - 2\delta_{i,2}\delta_{j,2})ab/p_{i,j} \tag{3.8}$$

$$= \langle a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2\rangle \tag{3.9}$$

However, this convention does not give us the minimum variance estimate of $\langle S(x)\rangle$ since the setting distribution $p_{i,j}$ is known. Next, we will introduce an improved data analysis protocol which is known as SD-based protocol. This protocol works under the IID assumption of the CHSH experiment.

3.2.2 Introduction to SD-based Protocol

As mentioned in the previous section that the estimate of $\langle S(x)\rangle$ using $\bar{S}$ is not ideal. Therefore, a new estimate $\tilde{S}$ is introduced and the expression of $\tilde{S}$ is given by:

$$\tilde{S} = \sum_{i,j} p_{i,j} \frac{\sum_{a,b} n(i,j,a,b) I(i,j,a,b)}{\sum_{a,b} n(i,j,a,b)} \tag{3.10}$$

$$= \sum_{i,j} \frac{\sum_{a,b} n(i,j,a,b)(1 - 2\delta_{i,2}\delta_{j,2})ab}{\sum_{a,b} n(i,j,a,b)} \tag{3.11}$$

$$= \sum_{i,j} (1 - 2\delta_{i,2}\delta_{j,2}) \langle a_i b_j \rangle \tag{3.12}$$

$$= \langle a_1 b_1 \rangle + \langle a_1 b_2 \rangle + \langle a_2 b_1 \rangle - \langle a_2 b_2 \rangle \tag{3.13}$$

where $n(i, j, a, b)$ is defined as the number of trials with the corresponding settings and outcomes of i, j, a and b. The above expression mimics the form of the original CHSH inequality which is essentially the "sum of means" in contrast of the "mean of sums" in the example of $\bar{S}$.

Subsequently, one has to find the standard deviation of $\tilde{S}$ so that the violation of local realism can be objectively quantified. In an event that a violation of local realism is observed (i.e. $\tilde{S} > B$), we can express the violation by standard deviations by computing $\frac{\tilde{S}-B}{\sigma}$, where σ is the standard deviation of $\tilde{S}$.

Assuming that the results are IID, the distribution of $\tilde{S}$ is normally distributed. Hence, we can compute the p-value as follows:

$$p^{SD} = Q\left(\frac{\tilde{S} - B}{\sigma}\right) \tag{3.14}$$

where $Q(x)$ is the Q-function which is essentially the integral of a standard normal function from x to ∞. (Recall that the standard normal function is a Gaussian function with mean of 0 and standard deviation of 1)

3.2.3 Limitations of SD-based Protocol

Additionally, a huge drawback of SD-based protocol is that it does not account for memory effect. The approach is to assume that every single trial and measurement is independent and identically distributed, which is not realistic in the experiments. In the next section, another protocol will be introduced to relax this assumption.

3.3 Relaxing IID: Martingale-based Protocol

In this section, a modification of the SD-based protocol is made to account for non-IID trials of the Bell experiment. This protocol is based on the Martingale theory which deals with random variable which behaviour based on the outcome of the previous runs, also known as "memory effect".

Here, the Martingale theory will be introduced and the framework of the Martingale-based protocol will be put forth.

3.3.1 Martingale

Before diving into the Martingale-based protocol, it is crucial to introduce the idea of Martingale. Consider a situation where repeated measurements will be made on a random variable X. On the kth trial, the measurement outcome will be denoted as X_k. Therefore, after n trials, a string of measurement outcomes $X_1, \cdots, X_n$ will be obtained. A Martingale has to fulfil the following condition:

$$\langle X_{n+1} | X_1, \cdots, X_n \rangle = X_n \tag{3.15}$$

By induction, the expectation value of X for every trial is the same, therefore, the knowledge of the outcomes from past trials do not give any advantage in predicting the outcomes of the future trials. However, Martingale is a subset of two larger classes of stochastic processes known as sub-martingale and super-martingale and are defined as follow:

$$\langle X_{n+1} | X_1, \cdots, X_n \rangle \geq X_n \text{ (sub-martingale)} \tag{3.16}$$

$$\langle X_{n+1} | X_1, \cdots, X_n \rangle \leq X_n \text{ (super-martingale)} \tag{3.17}$$

In these cases, the expectation values of X are not no longer guaranteed to be constant throughout the trials. Consider two parties making a guess of the next outcome of X, one of them has the knowledge of the outcomes of previous trials and the other do not have the information. Clearly, the one with the information of previous trials has an advantage other.

Relating the above exercise to Bell experiment, the expectation value of $S(x)$ measured in the experiment may depend on previous outcomes, also known as the presence of memory effect. In fact, the measurement made in a Bell experiment can be interpreted as a super-martingale, which will be proven in the next section.

3.3.2 Introduction to Martingale-based Protocol

Let us consider a super-martingale time sequence $M_k = \sum_{l=1}^{k}(S(x_l)-B)$. In this section, we assume that the measurement settings are chosen randomly and independently by both parties according to the distribution $p_{i,j}$. For all local realistic models, given the information of all settings and outcomes of previous trials, denoted by W_k, the expectation value of M_k is given by:

$$\langle M_k|W_k\rangle = \langle S(x_k) - B + M_{k-1}|W_k\rangle \tag{3.18}$$
$$= \langle S(x_k)|W_k\rangle - B + \langle M_{k-1}|W_k\rangle \tag{3.19}$$
$$= \langle S(x_k)|W_k\rangle - B + M_{k-1} \tag{3.20}$$
$$\leq M_{k-1} \tag{3.21}$$

In the above exercise, we invoked the identity of $\langle S(x_k)|W_k\rangle \leq B$ as for all local realistic models, regardless of any information of prior settings and outcomes, the Bell inequality will be satisfied. Hence, it is shown that the time sequence, M_k, is indeed a super-martingale.

Additionally, we know that the range of values of $S(x_k)$ is $-4 \leq S(x_k) \leq 4$. Hence, this implies that the increment between every subsequent term of M_k, mathematically written as $M_k - M_{k-1}$ is bounded by 2 extrema. The upper bound, b_u, is given by $4 - B$ while the lower bound, b_l, is given by $-4 - B$.

Finally, it is proven that for any local realistic model, the time sequence M_k is a super-martingale with bounded increment as shown above. Now, we can invoke the Azuma-Hoeffding inequality to construct the p-value of the Martingale-based protocol. For any local realistic model, the probability that an estimate $\hat{S}_{LR}$ is greater than or equals to the experimentally observed $\hat{S}$ is give by:

$$P_{LR}(\hat{S}_{LR} \geq \hat{S}) = P_{LR}(M_n \geq n(\hat{S} - B)) \tag{3.22}$$
$$\leq exp\left(-\frac{2n(\hat{S}-B)^2}{(b_u - b_l)^2}\right) \tag{3.23}$$

The above exercise has provided with an upper bound of $P_{LR}(\hat{S}_{LR} \geq \hat{S})$

which implies a valid p-value of:

$$p^{mart} = exp\left(-\frac{n(\hat{S} - B)^2}{16}\right) \tag{3.24}$$

In the above equation, we have substituted $b_u - b_l = 8$.

3.3.3 Limitations of Martingale-based Protocol

According to Zhang [15], the Martingale protocol is suboptimal, which means that more trials are required to observe the same level as of violation as compared to a protocol which is asymptotically optimal. We will discuss in details about the notion of this optimisation in the next section.

Additionally, the Martingale protocol is constructed based on a particular Bell inequality, and hence, posing a constraint on user who might wish to deal with other scenarios of Bell experiments.

3.4 Asymptotically Optimal: PBR Protocol

In this section, a protocol with an entire different approach will be introduced. The prediction based ratio (PBR) protocol exploits the statistical distance between the settings-outcomes frequencies and its closest distribution constraint by local realistic model to compute its p-value.

This protocol is said to be asymptotically optimal which meant that with a given amount of experimental data, the protocol will reflect its maximum amount of violation in the asymptotic limit.

3.4.1 The notion of Statistical Strength

As discussed in the earlier section, the goal of the data analysis of a Bell experiment is to find the p-value of the experimental result. Recall that if the p-value takes up a small numerical value, it would provide a strong evidence for the violation of local realism in the Bell experiment.

As more non-local data are collected and analysed, the stronger evidence it present against local realistic model and therefore p-value will decrease. The relationship between p-value and confidence gain rate, G, after obtaining n sets of data is defined as follows:

$$p = 2^{-Gn} \tag{3.25}$$

An optimal confidence gain rate is defined to be the statistical strength which will result in the most rapid decay of p-value. Hence, if a data analysis protocol is asymptotically optimal, it is able to discredit the null hypothesis with the least amount of data.

3.4.2 The Kullback-Leibler Divergence

The Kullback-Leibler Divergence (KL Divergence), also known as relative entropy in some literature, is a quantity which measures the difference between 2 different probability distribution. The KL Divergence from distribution q to p, denoted by $D_{KL}(q||p)$, is given by:

$$D_{KL}(q||p) = \sum_x q_x \log_2 \left[\frac{q_x}{p_x}\right] \tag{3.26}$$

Even though KL Divergence may represent some kind of statistical distance between distributions q and p but by definition, it is not a true metric due to its asymmetrical property (ie. $D_{KL}(q||p) \neq D_{KL}(p||q)$ in general).

The KL Divergence from q to p is the number of bits of information lost when p is approximated to q. In its application in hypothesis testing, given the experimental data, x, the data analyst will like to find out if probability distribution p can be responsible for producing x, where in fact, q is the "true" probability distribution of the data. A large KL Divergence will provide strong evidence against the hypothesis that probability distribution p generates the data x. Given non-zero prior probabilities on q and p, the posterior probability that p generates x rather than q is given by $2^{-nD_{KL}(q||p)+O(n)}$ [13].

By defining the appropriate probability distributions to be q and p, the KL divergence can be interpreted as the statistical strength of a hypothesis testing [13]. In fact, the optimal confidence-gain rate for rejecting p in favour of q is given by $D_{KL}(q||p)$.

3.4.3 Introduction to the PBR Protocol

The Prediction Based Ratio (PBR) protocol was constructed to address the issues of memory effect of the SD-based protocol and the sub-optimality of the Martingale-based protocol. In this protocol, we will rewrite the Bell inequality and we will prove that this method of proving the violation of

local realism is asymptotically optimal.

First, we define the PBR, $R_k(x)$, given settings and outcomes, x, for the k^{th} trial to be:

$$R_k(x) = \frac{q_x^{(k)}}{p_{LR,x}^{(k)}} \tag{3.27}$$

where $q_x^{(k)}$ is defined to be an estimate of the k^{th} setting and outcome combinations distribution based on information prior to the k^{th} trial of the experiment. In the next section, possible methods that can be used to estimate future setting and outcome combinations will be discussed. $p_{LR,x}^{(k)}$ is the probability distribution of setting and outcome combinations constrained by local realistic models such that the KL Divergence from $q_x^{(k)}$ to $p_{LR,x}^{(k)}$ is the minimum over all sets of distribution allowed by local realistic model.

The PBR, $R_k(x)$, is constructed such that it is non-negative and for any local realistic model, $\langle R_k(x) \rangle \leq 1$ given that the setting distribution is given by $p_{i,j}$. This inequality condition is the Bell inequality in the case when PBR protocol is used to analyse Bell experiment data because it defines the upper bound of the quantity $\langle R_k(x) \rangle$ which can be accounted for by local realistic model. In order to take into account of any possible memory effect in the experiment, we introduce a new variable P_k, and it is defined by:

$$P_k(x) = \prod_{l=1}^{k} R_{l-1}(x_l) \tag{3.28}$$

Similar to the Martingale protocol, one is able to find the conditional expectation of P_k given all possible available information of the setting and outcome combinations before the k^{th} trial, denoted by W_k. It follows that the conditional expectation of P_k for any local realistic model is given as follows:

$$\langle P_k|W_k\rangle = \left\langle \prod_{l=1}^{k} R_{l-1}(x_l)|W_k \right\rangle \tag{3.29}$$

$$= \left\langle \prod_{l=1}^{k-1} R_{l-1}(x_l) \times R_{k-1}(x_k)|W_k \right\rangle \tag{3.30}$$

$$= \langle P_{k-1} \times R_{k-1}(x_k)|W_k\rangle \tag{3.31}$$

$$= P_{k-1} \times \langle R_{k-1}(x_k)|W_k\rangle \tag{3.32}$$

$$\leq P_{k-1} \tag{3.33}$$

In the fourth step of the proof above, notice that the conditional expectation of P_{k-1} given W_k is given by P_{k-1} itself because with all the information about the setting and outcome combinations prior to k^{th} trial, we can determine precisely the value of P_{k-1}. In the final step of the proof above, the inequality is a direct consequence of the previous claim that for any local realistic model, $\langle R_k(x)\rangle \leq 1$.

Hence, it is proven that for any local realistic model that $\langle P_k\rangle \leq \langle P_{k-1}\rangle$, one can conclude that for any local realistic model, the variable P_k is indeed a supermartingale. Additionally, notice that $P_1 = R_0$ and since $\langle R_k(x)\rangle \leq 1$, it implies that $\langle P_1\rangle \leq 1$. Using the relation $\langle P_k\rangle \leq \langle P_{k-1}\rangle$, we can conclude that for any local realistic model, $\langle P_k\rangle \leq 1$ for all k.

Notice that unlike in the Martingale protocol, the values of P_k ranges from 0 to ∞ and hence, there are no finite bounds on the values of P_k. Even though P_k is a supermartingale, Azuma-Hoeffding Inequality is incompatible with the PBR protocol and an alternative method has to be used to find the p-value.

3.4.4 P-value of the PBR Protocol

Similar to previous protocols, suppose that the Bell experiment is run for n trials and the experiment results $x_1, \cdots, x_n$ are obtained. Using these results, we are able to obtain a specific value for P_n, which will be denoted by $\hat{P}$. Hence, it will be reasonable to define the p-value as the probability of $P_n \geq \hat{P}$ given by any local realistic model. Hence, we are able to exploit the statistical nature of P_n to find a simpler form for the p-value.

Let us introduce an indicator random variable, I_P such that:

$$I_P = \begin{cases} 1 & \text{if } P_n \geq \hat{P} \\ 0 & \text{if } P_n < \hat{P} \end{cases}$$

Notice that $\hat{P}$ is non-negative as earlier defined. Therefore, if one multiplies P_n by I_P, notice that if $P_n < \hat{P}$, the product will be simply 0 but if $P_n \geq \hat{P}$, the product will be simply given by P_n which has a value smaller than $\hat{P}$. In both cases, we have similar conclusion and we can safely write down the following inequalities:

$$\hat{P} \times I_P \leq P_n \tag{3.34}$$

$$\implies \langle P_n \rangle \geq \left\langle \hat{P} \times I_P \right\rangle \tag{3.35}$$

$$\geq \hat{P} \times (1 \times \Pr(P_n \geq \hat{P}) + 0 \times \Pr(P_n < \hat{P})) \tag{3.36}$$

$$\geq \hat{P} \times \Pr(P_n \geq \hat{P}) \tag{3.37}$$

The above inequality can be rearranged into:

$$\Pr(P_n \geq \hat{P}) \leq \frac{\langle P_n \rangle}{\hat{P}} \tag{3.38}$$

The above inequality relation is also known as the Markov inequality. Previously we have established that for any local realistic model, $\langle P_n \rangle \leq 1$. However, in the case that $\langle P_n \rangle \geq \hat{P}$, we have a value for probability greater than 1, which is unacceptable. Therefore, in these cases, we set probability to 1. Therefore, the probability of $P_n \geq \hat{P}$ given by any local realistic model is given by:

$$\Pr_{\text{LR}}(P_n \geq \hat{P}) \leq \min(\frac{1}{\hat{P}}, 1) \tag{3.39}$$

Apparently, using the method above, it is not possible to obtain a tight value of the p-value. However, an observation of a small valued upper bound of the p-value is sufficient to reject the null hypothesis. Alas, we define the p-value of the PBR protocol to be:

$$p^{\text{PBR}} = \min(\frac{1}{\hat{P}}, 1) \tag{3.40}$$

3.4.5 Estimating settings-outcomes distribution

In order to apply the PBR protocol on any Bell experiment results, it is essential to have a standard procedure to predict future settings-outcomes distribution. The approach in PBR protocol is to find an estimation q' of the "true" probability distribution q of the settings-outcomes combinations.

After n trials of the Bell experiment, the settings and outcomes $x_1, \cdots, x_n$ will be known. Without the knowledge of the state of the measured particles, the only information that the prediction of q can only be based on the empirical frequencies, f_x, given by:

$$f_x = \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k,x} \tag{3.41}$$

The 2 constraints that has to be satisfied by q and likewise q' are as follows:

1. No-signalling constraint: $P(a|i,j) = P(a|i)$ and $P(b|i,j) = P(b|j)$.
2. Setting distribution constraint: Since the setting distribution, $p_{i,j}$ is known, q has to fulfil the following equation: $p_{i,j} = \sum_{a,b} q_{i,j,a,b}$

Now, we denote the set of probability distributions which satisfy the above constraints to be V. By assuming IID trials, we are able to compute a first estimate of q, denoted by q_0, as follows:

$$q_0 = \arg\max_{q' \in V} L(f|q') \tag{3.42}$$

$$\text{where } L(f|q') = \prod_x q_x^{nf_x} \tag{3.43}$$

The above is the Maximum Likelihood Estimation problem which can be solved numerically using the sequential quadratic programming. This algorithm is designed to solve non-linear optimisation problem with linear constraints.

There are 2 main problems foreseen to arise if q_0 computed is to be used to compute the p-value from PBR protocol. The first problem arises when there exist zero probabilities in some settings-outcomes combination in q_0 which will cause the resulting p-value to be 1 with no chance of recovery. A

simple solution would be to incorporate an uniform distribution, u, into the estimated probability distribution as follows:

$$q_1 = \frac{n}{n+1} q_0 + \frac{1}{n+1} u \tag{3.44}$$

The second problem involves the initial learning transient as a result of insufficient information of the system to make useful estimates of future settings and outcomes. If the states of the measured particles are known, it is possible to estimate the probability distribution even before the experiment which solves the initial learning transient problem.

However, in many problems, the state of the particles are usually unknown prior to the measurement. In these situations, alternative solutions are required to solve the initial learning transient problem. A solution is to set R_k to be 1 until a substantial Bell violation had been observed. Additionally, the protocol should be run in blocks of experimental data so that in each "run", there is sufficient information about the experiment to give an optimal p-value.

In order for the PBR protocol to adapt to any changes in the experiment, the estimation of future settings and outcomes should be based on experiment data of the recent trials. This can be achieved by making adjustment of the data half-life, λ_d.

3.4.6 Proof of asymptotical optimality

Taking the asymptotic case of a Bell violating IID experiment, the estimated probability distribution, $q_x^{(k)}$ and the frequency of the settings-outcomes, f_x, will converge to the true settings-outcomes distribution, q. Similarly, $p_{LR,x}^{(k)}$ will converge to the corresponding p_{LR}, optimal to q. It can be shown that PBR protocol is indeed asymptotically optimal by the following argument.

Consider the $-\log p$ increment given the experimental data of an IID Bell experiment:

$$-\log p \text{ increment} = \sum_x f_x \log \frac{q_x^{(k)}}{p_{LR}^{(k)}} \tag{3.45}$$

$$= \sum_x q_x \log \frac{q}{p_{LR}} \tag{3.46}$$

$$= D_{KL}(q||p_{LR}) \tag{3.47}$$

Notice that the expected $-\log p$ increment of an IID Bell experiment reduces itself to the KL divergence from the true settings-outcomes probability distribution to the optimal settings-outcomes distribution constraint by LR. As mentioned in the previous section, the KL divergence shows the amount of evidence against the hypothesis that p_{LR} generates the experimental data.

Since in the IID limit, $-\log p$ converges to the $D_{KL}(q_x||p_{LR})$ which is the statistical strength of the hypothesis testing against local realism, then the p-value generated by the PBR protocol is asymptotically optimal.

3.5 Geometric Interpretation of Hypothesis testing

After much discussions on hypothesis testing, this section seek to describe hypothesis testing on Bell violation using geometrical interpretation. In hypothesis testing mentioned in previous sections, the aim is to find an estimate of the settings-outcomes probability distribution and determine if the estimate falls within the LR polytope.

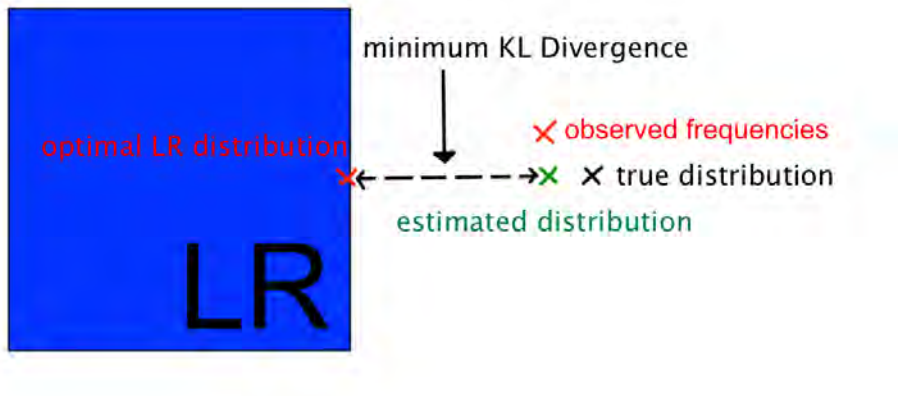

Figure 3.1: PBR protocol in geometrical representation. The blue square denoted "LR" represents the LR polytope which contains all probability distributions which fulfils the local realism constraints. The black cross represents the true probability distribution while the aim is to find its best estimate (green cross) with the information of the observed frequencies (red cross). Upon finding the estimate distribution, the optimal LR distribution can be computed which is required by the protocol.

Figure 3.1 shows the framework of the PBR protocol using geometric

representation as discussed earlier on. In the figure, the blue square represents the LR polytope where all probability distributions that can be accounted for by local realism models reside in. The facets of the LR polytope represents the Bell inequalities as mentioned earlier on.

In the PBR protocol, using the observed settings-outcomes frequencies generated by the Bell experiment (represented by the red cross), the data analyst is to find an estimation (represented by the green cross) of the true probability distribution (represented by the black cross) of the system. Using the estimate, the analyst will then be able to find an optimal LR distribution (represented by the red cross in the LR polytope) by minimising the KL Divergence of the estimate with any point in the LR polytope. Finally, the p-value can be computed using the estimated distribution and the optimal LR distribution as discussed in the previous section.

Chapter 4

Results

Since there exist different data analysis protocols with varying relaxation of assumptions on the Bell experiment, a comparison between these protocols should be made via implementation of these protocols on some concrete situations where complete information about the experiment is known. In the following non-IID situations, it will underline the importance of having a protocol which relaxes the IID assumption to avoid drawing the wrong conclusion from a set of Bell experiment results.

Thereafter, the aim is to have a protocol which also enables data analyst to conclude the amount of private randomness that is generated from the Bell experiment. An attempt was made with promising results which can potentially open the doors to an avenue to quantify randomness from non-IID Bell experiment.

4.1 Non-IID Situations

In order to test the robustness of the data analysis protocols mentioned in the earlier chapters, one has to consider Bell experiment with non-IID states. In this section, non-IID situations known as the "2-days experiments" will be introduced and data analysis protocols introduced in the last chapter will be used in these experiments.

4.1.1 Situation A $(2\sqrt{2}\ ,\ 2)$

Consider a situation where by the data collected from the Bell experiment is produced by maximally entangled state which gives maximal Bell vio-

lation(CHSH= $2\sqrt{2}$). However, on the second day of the experiment, the source of the maximally entangled particles is removed and the Bell experiment produces deterministic outcomes (CHSH=2). In this situation, some of the outcomes are non-local while some are not.

If one picks up the settings and outcomes data produced by the Bell experiment and assumes that the outcomes produced are IID, then one will conclude that the CHSH value will be in-between 2 and $2\sqrt{2}$, depending on the amount of data produced in each day. Recall that for SD-based protocol, the p-value depends on the difference between the observed value of CHSH and 2. In the case where by the amount of data produced on day 2 is much larger than day 1, the CHSH value will approach to 2. The above discussion demonstrates that by using protocols which assumes IID trials in non-IID situation, the presence of random outcomes diluted in large amount of deterministic data will not be detected.

On the other hand, the PBR protocol is able to recognise the presence of non-local correlations due to the small but finite KL divergence from $q_x^{(k)}$ to $p_{LR,x}^{(k)}$. Therefore, the PBR protocol is able to provide a strong evidence against local realism despite the size of the non-local settings-outcomes data is small with respect to that of the local settings-outcomes data.

Figure 4.1 shows a graph plotted $-\log_2 p$ value against the number of trials resulted by performing SD-based protocol and PBR protocol on the same set of data produced by the 2-day experiment(Situation A). In this paper, the results of any data analysis protocol on Bell experiment will be presented in $-\log_2 p$, instead of p-value, since the resultant p-values in this paper vary over orders of magnitude.

Referring to Figure 4.1, the responses to the deterministic data on day 2 by the two protocols are clearly distinct. The $-\log p$ value given by the SD-based protocol decays rapidly to 0 but the $-\log p$ value given by the PBR protocol decays and gradually stabilise at 70% of its maximum value. It is apparent that two opposite conclusions are drawn from the two different data analysis protocols. The PBR based protocol confirms that the data from the experiment cannot be explained by any local realistic model but the SD-based protocol suggests the otherwise.

This experiment explicitly shows that the SD-based protocol is unable to cope with the change in state of the experiment and hence, giving a false interpretation of the system.

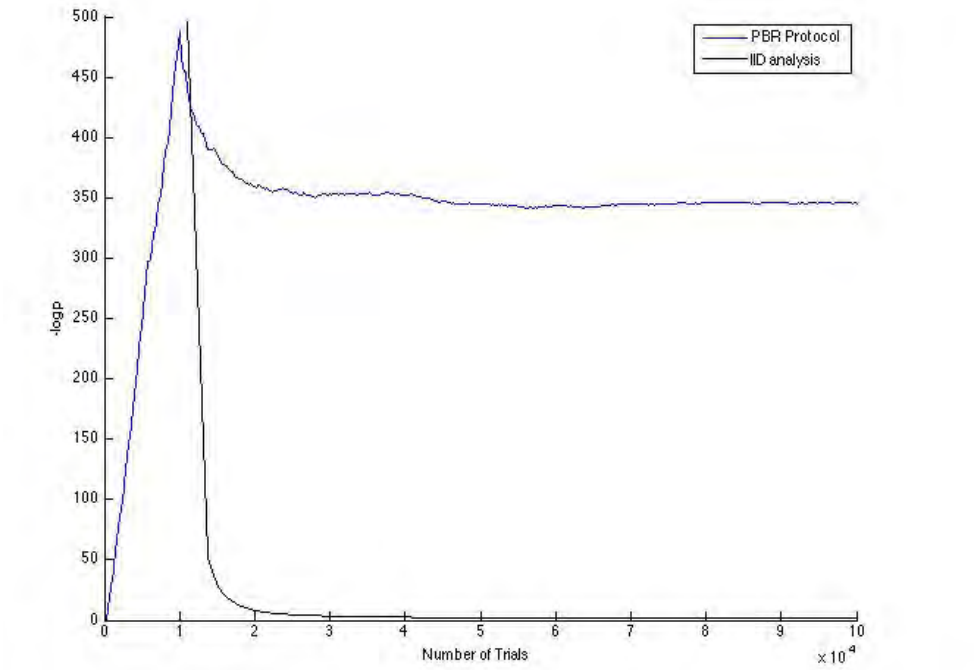

Figure 4.1: This is a graph of the computed $-\log p$ against number of trials for the 2-days experiment where the first 10% of the data has a maximum Bell violation (i.e. CHSH=$2\sqrt{2}$) where is subsequent 90% of the data is deterministic (i.e. CHSH = 2) using the IID analysis (SD-based protocol) and PBR protocol. For the IID analysis results, the initial p-value is numerically smaller than the accuracy of MATLAB but in the region of local data, the $-\log p$ value plunges to 0 rapidly. For the PBR protocol, the graph shows that in the region of deterministic data, the decay of the graph slows down quickly and stabilises.

4.1.2 Situation B ($2\sqrt{2}$, $-2\sqrt{2}$)

In this section, a variant of the "2-days experiment" will be examined. In this situation, instead of producing deterministic outcomes on day 2, the state of the source is changed to that of the opposite maximum Bell violation (CHSH=$-2\sqrt{2}$). Clearly in this situation, Bell inequality is violated throughout the experiment, and it is known that all the outcomes produced by this experiment are non-local.

Here, an ideal data analysis protocol will be able to account for the change in state of the system and recognise that the outcomes produced are all non-local regardless of the relative size of the data produced in day 1 and day 2.

Recall that the choice of $q_x^{(k)}$ in the PBR protocol is free and the conventional method used is mainly based on the frequencies of the setting-outcome

combinations. The conventional method of estimating $q_x^{(k)}$ may not be ideal in situation B because correlated outcomes between Alice and Bob in day 1 will be anti-correlated in day 2 and anti-correlated outcomes in day 1 will be correlated in day 2. This means that the overall frequencies in the "2-days experiment" will reflect a statistics of uncorrelated outcomes between Alice and Bob. Thus, by estimating $q_x^{(k)}$ via the conventional means, such $q_x^{(k)}$ can be explained by a local realistic model and hence a $-\log p$ value of 0.

Since the choice of $q_x^{(k)}$ for a given Bell experiment data is free, it is allowed to let the influence of the data obtained on the estimation of $q_x^{(k)}$ decay with a specific data half-life, denoted by λ_d. In the previous case of the "2-days experiment" (Situation A) the data half-life, λ_d, is set to be infinite. This corresponds to the situation where by all the data from every trials used in the analysis protocol is given equal weightage. For any finite data half-life, the data obtained from the experiment is "forgotten" at a certain rate during the computation of $q_x^{(k)}$.

As mentioned earlier, the PBR protocol analyse data in blocks with the frequencies of settings-outcomes events in the k^{th} block is denoted by b_k. Also, the size of block is denoted by N_B. Finally, the expression for the frequencies of settings-outcomes events used for computation of $q_x^{(k)}$ after $(N_B * k)^{th}$ trials, denoted by f^k, is given by:

$$f^k = (1 - x_k)f^{k-1} + (x_k)b_k \tag{4.1}$$

where

$$x_{k+1} = \frac{2^{\frac{N_B}{\lambda_d}} x_k}{1 + 2^{\frac{N_B}{\lambda_d}} x_k} \ , \ x_1 = 1 \tag{4.2}$$

The above expressions show that by varying λ_d, it is possible to obtain different but equally valid settings-outcomes distribution estimation, $q_x^{(k)}$. Figure 4.2 shows a graph plotted $-\log p$-value against number of trials of the "2-days experiment" with varying λ_d.

As seen in Figure 4.2, the graph corresponding to $\lambda_d = \infty$ shows a significantly lower $-\log p$ value as compared to cases where λ_d is finite. Also, it is observed that with 10,000 trials conducted in day 1, only 1,700 trials that are conducted in day 2 are required to bring down the $-\log p$-value to 0 and stabilises. Subsequently, after conducting 61,200 trials in day 2, the graph will begin to pick up a positive gradient.

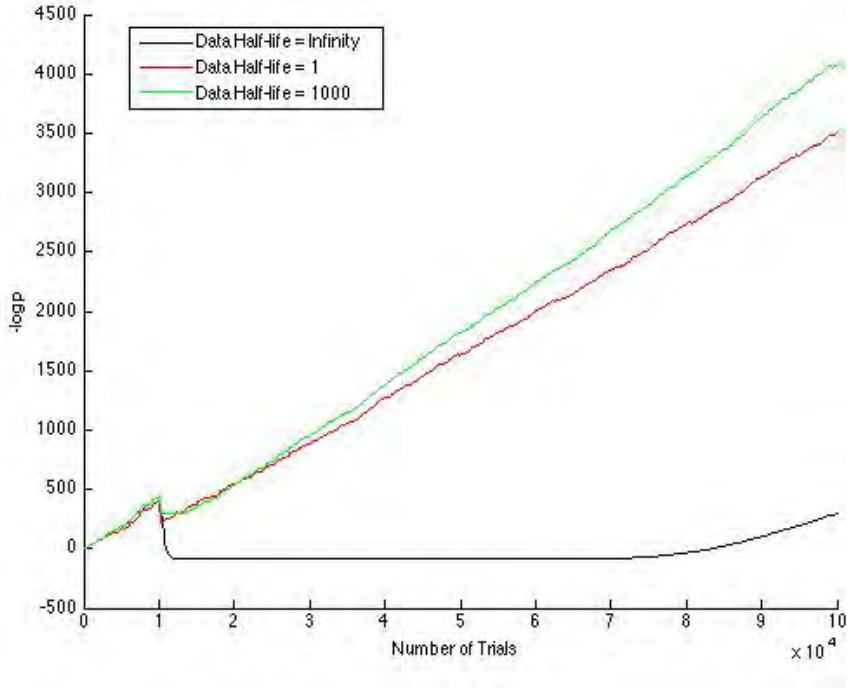

Figure 4.2: This is a graph of the computed $-\log p$ against number of trials with the first 10% of the data having CHSH value of $2\sqrt{2}$ and the subsequent 90% of the data having CHSH value of $-2\sqrt{2}$ using the PBR protocol. A series of analysis had been processed with varying values of data half-life and it is found that the data-half life of 1000 gives an optimisation of $-\log p$ value given that the block size of 200. When the data half-life is smaller than 1000, the gradient of the graph is sub-optimal. However, when the data half-life exceeds 1000, the observed 'dip' at day 2 will be pronounced.

This unusual phenomenon occurs because the settings-outcomes frequencies, f, is the mixture of 2 different sets of settings-outcome frequencies produced by sources that violates Bell's inequality in the opposite directions. The proportion of the individual constituent sets of settings-outcome frequencies will define the properties of f and in turn, have an effect on the distribution estimation, $q_x^{(k)}$. Thus, such changes will be reflected in the $-\log p$ value processed by the PBR protocol.

In order to explore the behaviour of the the overall settings-outcomes frequencies, f, it is intuitive to write down the expression for the CHSH of f as a function of the proportion, β, of the settings-outcome frequencies of day 2, which is given by:

$$\text{CHSH} = 2\sqrt{2}(1-\beta) - 2\sqrt{2}\beta \tag{4.3}$$

As mentioned earlier, for values of β between $\frac{17}{117}$ and $\frac{612}{712}$, the PBR protocol gives a $-\log p$-value of 0. This range of values of π corresponds to the following inequality $-2.03 < \text{CHSH} < 2.00$. The result is not surprising because the statistics of the overall frequencies can be explained using local models, even though it does not truly reflect the actual situation of the experiment.

On the other hand, if the data half-life, λ_d, is given a finite value, the data produced by the Bell experiment at a later time will have a heavier weightage on the computation of f. Referring to Figure 4.2 again, the graph corresponding to $\lambda_d = 1$ and $\lambda_d = 1000$ show a similar dip in $-\log p$ value as the case of infinite data half-life when the experimental data from day 2 is introduced. However, for cases of finite data half-life, the $-\log p$ value increases subsequently because the influence of data from day 1 on the computation of $q_x^{(k)}$ becomes negligible.

By varying λ_d and obtaining its corresponding $-\log p$ value, it is observed that the optimal data half-life that produces the highest $-\log p$ value is $\lambda_d = 1000$. There are 2 competing effects in play here: if λ_d has a large value, then there is a large proportion of day 1 data used in the computation of $q_x^{(k)}$, lowering the $-\log p$ value. On the other hand, if λ_d has a small value, then there will be a higher fluctuation of the frequencies giving rise to a large statistical distance between the frequencies and the "true" settings-outcomes distribution, thus, lowering the $-\log p$ value.

4.1.3 Reviewing Situation A

In the previous section, it was concluded that it is essential to set a finite data half-life, λ_d, for non-IID Bell experiments, in order to obtain the right conclusion from the PBR protocol.

Additionally, recall that in Situation A, while an infinite λ_d does not render the conclusion given by PBR protocol invalid, it causes a "dip" in the $-\log p$-value. This is undesirable because additional deterministic outcomes on day 2 should not affect the extent of Bell violation observed.

Figure 4.3 shows the analysis generated by PBR protocol done with different λ_d with similar conclusion. However, PBR protocol with a finite λ_d provides users with a p-value that is representative of the extent of Bell violation observed in the Bell experiment.

Initially, a "dip" in $-\log p$-value with infinite λ_d could imply that there

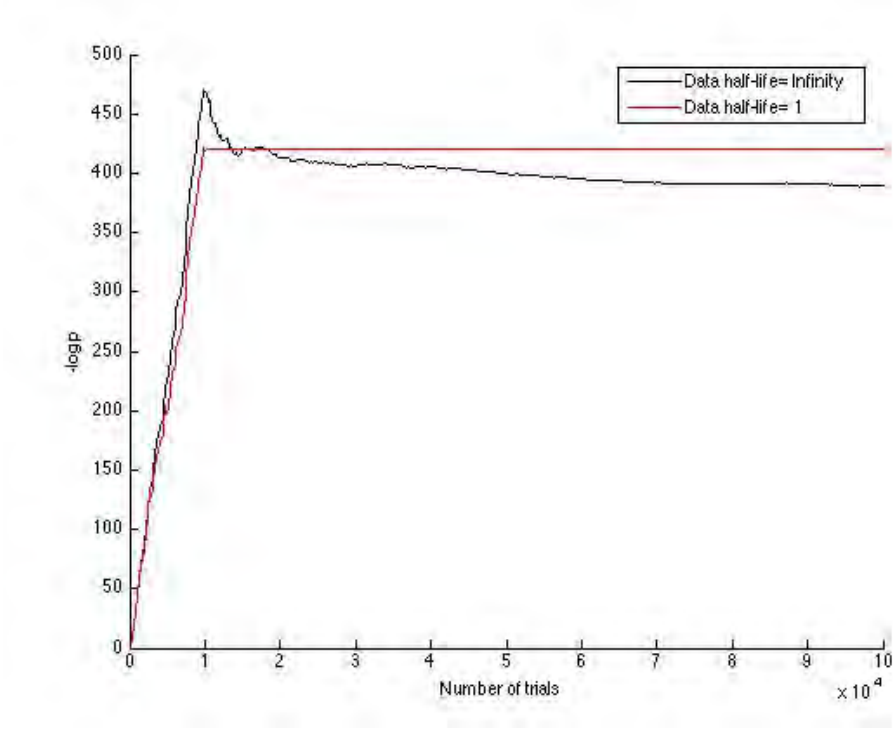

Figure 4.3: The graph plotted $-\log p$ against number of trials of Situation A with varying data half-life. The graph shows that with a finite λ_d, the "dip" observed initially can be prevented.

exist deterministic or oppositely Bell violating data. However, with finite λ_d, deterministic data has no observable effect on $-\log p$-value. Hence, by inspection of the graph, data analysts can interpret the nature of the experimental data.

4.2 Randomness from PBR protocol

The PBR protocol was designed by Zhang and his collaborators to falsify local realism given the experimental data from a Bell experiment with possible memory effect. Hence, the PBR protocol does not provide any avenue to quantify the amount of randomness produced from Bell violation.

The hope in this section is to provide a stepping stone for any future ambition to quantify randomness from non-IID Bell experiment. This section will illustrate that by making modification to the PBR protocol, it is possible to determine a lower bound of randomness obtained from non-IID Bell experiment outcomes.

4.2.1 Changing the Null Hypothesis

Recall that previously, the null hypothesis for all the above mentioned protocols is defined as the results produced by the Bell experiment cannot be explained by any local realistic model. Here, it is more useful to phrase the null hypothesis in terms of Bell violation. For the examples in the previous chapters, we have the hypotheses:

H_0 : All trials have values of $CHSH$ less than or equals to 2

H_a : There exists trials with value of $CHSH$ greater than 2

The above hypotheses describes a hypothesis testing that gives 2 types of conclusion, either the experimental data gives statistical evidence to prove or disprove the existence of any experimental outcomes which violates the Bell inequality, CHSH ≤ 2. Notice that by varying the null hypothesis and increase the number of hypothesis tests on the same experimental data, we are able to probe the "strength" of Bell violation using a series of these hypothesis tests. Each hypothesis test can be represented by just its null hypothesis since the alternative hypothesis will then be implied.

In order to illustrate this idea, a series of null hypotheses are constructed below:

$H_{0,0}$: All trials have value of $CHSH$ less than or equals to 2

$H_{0,1}$: All trials have value of $CHSH$ less than or equals to 2.1

$H_{0,2}$: All trials have value of $CHSH$ less than or equals to 2.2

$H_{0,3}$: All trials have value of $CHSH$ less than or equals to 2.3

$H_{0,4}$: All trials have value of $CHSH$ less than or equals to 2.4

$H_{0,5}$: All trials have value of $CHSH$ less than or equals to 2.5

$H_{0,6}$: All trials have value of $CHSH$ less than or equals to 2.6

$H_{0,7}$: All trials have value of $CHSH$ less than or equals to 2.7

$H_{0,8}$: All trials have value of $CHSH$ less than or equals to 2.8

After running through the data analysis processes with each of these null hypotheses, the analysis will each generate a p-value. The p-values will increase as we vary from $H_{0,0}$ to $H_{0,8}$ and this can be explained by the following geometric interpretation. For simplicity, the CHSH value being tested

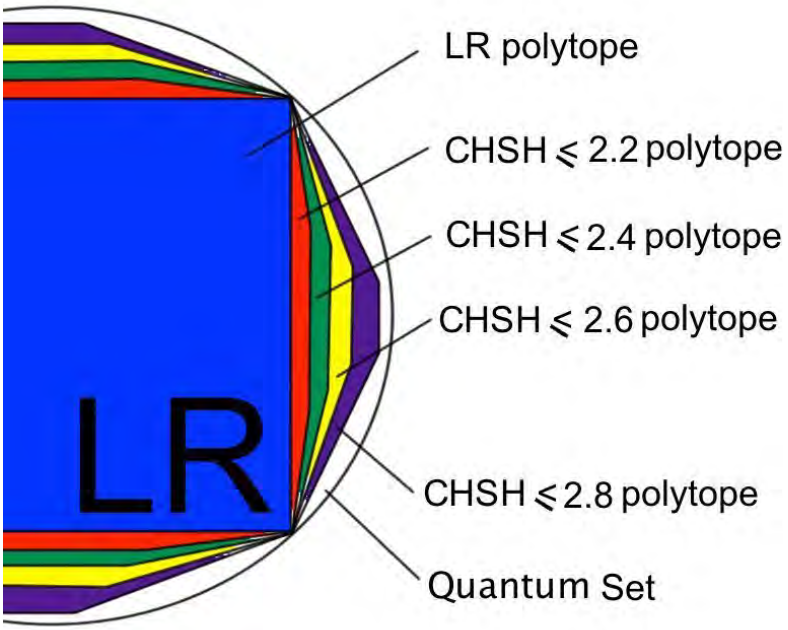

Figure 4.4: The expanding "test" polytope in geometrical representation. This figure shows that as the "test" polytope expands, it fills up the space which contains all distributions which are physically allowed. Hence, any position in space will have a reduced distance from the expanding "test" polytope.

against in a particular hypothesis test shall be defined as the hypothesis CHSH. (i.e. $H_{0,0}$ has a hypothesis CHSH to be 2)

The set of probability distributions that is fulfil by each null hypothesis can be described by a polytope. Now, we term the polytope which the estimated probability distribution, $q_x^{(k)}$, is tested against to be "test" polytope. As the "test" polytope expands, the proximity from any allowed position to the "test" polytope decreases and hence, the p-value increases.

The expansion of the "test" polytope can be visualised with the representation presented in figure 4.4. Even though the diagram is not a stereographic projection of the polytope on the 2-dimensional space, but it provides a good intuition of the evolution of the "test" polytope of the proposed scheme. As presented in the diagram, as the "test" polytope expands, the proximity of any point in the quantum set will decrease until it is being consumed by the polytope.

In order to find out the approximate position of the system's probability distribution, the modified hypothesis tests should run sequentially from $H_{0,0}$

to $H_{0,8}$. In this way, when the $-\log p$-value hits 0 for the first time in a particular test, the position of the estimated distribution, $q_x^{(k)}$, will be closest to the facet of the corresponding polytope than the other polytopes. This provides an intuition of the position of the estimated probability distribution, $q_x^{(k)}$.

Recall that in the preliminary chapter, it is established that the amount of randomness in the measurement outcome is related to the CHSH value. Therefore, the position of the estimated probability distribution will shed light on the amount of randomness present in the experimental outcomes analysed by the modified PBR protocol.

4.2.2 The Implementation of the Modification

In the PBR protocol, the p-value generated from a set of experimental result depends on the estimation, $q_x^{(k)}$ and the optimal LR distribution, $p_{LR,x}^{(k)}$. However, when the null hypothesis is changed, the usage of $p_{LR,x}^{(k)}$ will no longer be valid. Instead, an optimal probability distribution, denoted by p_{op} contained by the corresponding polytope should be considered.

$$p_{op} = \underset{p \in \text{test polytope}}{\arg\min} \; D_{KL}(q||p) \tag{4.4}$$

In order to obtain p_{op}, the above minimisation problem has to be solved. The most efficient approach is to study how $p_{LR,x}^{(k)}$ is obtained.

Finding $p_{LR,x}^{(k)}$

The first task will naturally be to identify the extremal points of the LR polytope as these points will effectively define the points which are included in the polytope. Also, these extremal points of the polytope can be interpreted as local vectors, $P(ab|ij\lambda)$, where each local strategy is represented by a λ value.

The local vector will define the allowed and forbidden correlations obtained in the Bell experiment by local realistic models. Given a set of experimental data generated by a Bell experiment which does not violate any Bell inequality, one can interpret the data as a result of a linear combination of these local strategies, λ.

The local vector is expressed in terms of a 16-by-16 matrix with row rep-

resenting each local strategy, λ and the column representing each settings-outcomes combination, $x = a, b, i, j$. Recall that $p^{(k)}_{LR,x}$ is the optimal probability distribution of settings and outcomes constraint by local realistic model and is essentially $P_{LR}(ab|ij)$. Hence, $p^{(k)}_{LR,x}$ can be written as:

$$p^{(k)}_{LR,x} = \sum_{\lambda} \rho(\lambda) P(ab|ij\lambda) \tag{4.5}$$

Where $\rho(\lambda)$ is the distribution of local strategies adapted. The above equation shows that $p^{(k)}_{LR,x}$ can be obtained by solving for $\rho(\lambda)$ since the local strategies, $P(ab|ij\lambda)$, are known.

For convenience, the following calculations to obtain $\rho(\lambda)$ will be written in matrix form and the indices i, j and k are matrix indices which should not be confused with the Bell experiment settings i and j. Now, we rewritten the local strategies in matrix form to be h_{ij} where indices $i, j \in [1, 16]$. Rewritting the minimisation problem:

$$\rho(\lambda) = \underset{f \geq 0, \sum f = 1}{\arg\min} \; D_{KL}(\{q_j\} || \{\sum_i f_i h_{ij}\}) \tag{4.6}$$

$$= \underset{f \geq 0, \sum f = 1}{\arg\min} \; \{\sum_j q_j \log_2(\sum_i f_i h_{ij})\} \tag{4.7}$$

These class of problems are well studied and its solution can be achieved via the Expectation Maximisation (EM) algorithm [14] as given below:

$$f^{(n)}_i = f^{(n-1)}_i \sum_j \left(\frac{h_{ij}}{\sum_k f^{(n-1)}_k h_{kj}} \right) q_j \ , \text{ for } n = 1, 2, \cdots \tag{4.8}$$

The above algorithm is to be iterated with any $f^{(0)}_i > 0$ and the numerical values of the array, f_i, will converge to $\rho(\lambda)$.

Finding p_{op}

As the "test" polytope expands out of the LR polytope in the proposed scheme, the extremal points of the "test" polytope changes. In this section, the main task is to find all extremal points of the "test" polytope as that will provide the full list of strategies, $h_{i,j}$, available within the polytope. Once, $h_{i,j}$ is obtained, p_{op} can be easily determined by using the above EM algorithm.

In our approach, the CHSH value of the "test" polytope is first identified in order to define the inequality constraints on the probabilities distributions. The Polyhedral Representation Transformation Algorithm (PORTA) [4] has the capabilities of interchanging maximal points and inequality constraints. In particular, the $traf$ function converts ieq files, which contains the list of inequalities and equalities constraints, to poi files, which contains list of maximal points, and vice versa. The list of 23 inequalities bounding the "test" polytope can be found in the appendix. After applying the traf function on the inequalities constraint, a total of 80 maximal points will be given. Hence, p_{op} can be obtained and used to test experimental data against an expanded polytope.

4.2.3 Results of the Modification

To prove the validity of the modified PBR protocol, the protocol is applied to simulated experimental results with known parameters. An agreement between the known parameters and the conclusion drawn from the modified PBR analysis will demonstrate its validity.

The analysis results of two different IID Bell experiments done using the modified PBR protocol are presented and discussed below. Finally, this discussion will be wrapped up by an attempt on quantifying private randomness from Bell experiment using the modified PBR protocol while relaxing the IID assumption.

Experiment A: IID Maximum Violating Bell Experiment

The experimental outcomes of a IID maximally violating Bell experiment (CHSH $= 2\sqrt{2}$) are used to run the modified PBR protocol. From previous experiments, a positive linear relationship between $-\log p$-value and the number of trials is expected of an experimental data of IID Bell experiment.

In figure 4.5, the linear relationship between $-\log p$ and number of trials is observed for all 9 hypothesis tests, in agreement with above prediction. Also, note that the final $-\log p$-value are positive values for all hypothesis tests as the "test" polytope expands. This implies that there exist some measurement outcomes which display non-local correlations with CHSH value greater than 2.8. Indeed, having to know that the true value of CHSH of the measured system is $2\sqrt{2}$, it confirms the findings of the modified protocol.

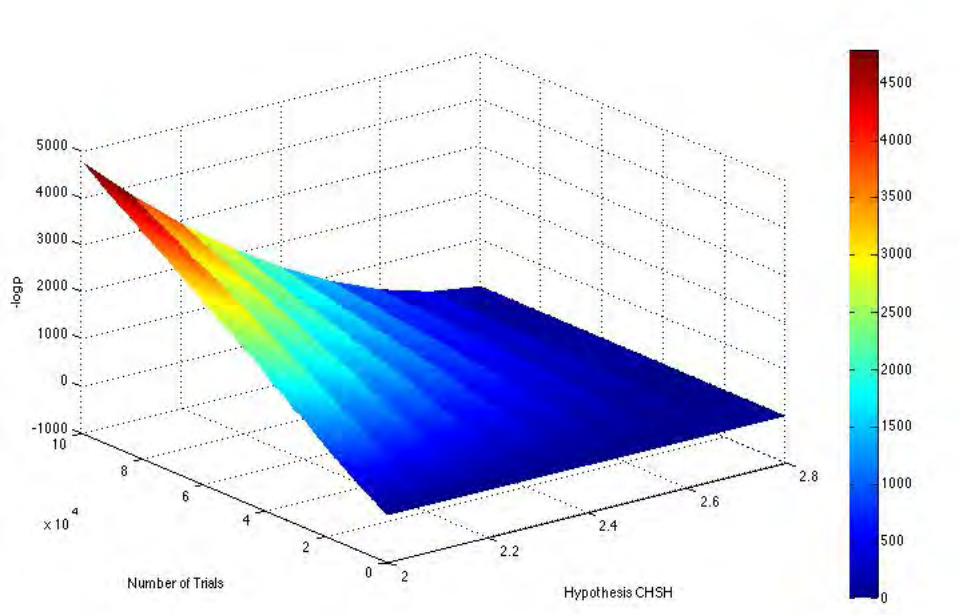

Figure 4.5: The surface plot of $-\log p$-values against number of trials against the hypothesis CHSH. The surface plot is generated by running the experimental data of an IID Bell experiment with maximum Bell violation on the modified PBR protocol.

Experiment B: IID Bell Experiment with CHSH=2.42

Now, experiment B aims to investigate the validity of the modified protocol when the CHSH of the experimental data is lower than the hypothesis CHSH. In this experiment, simulated experimental data from IID Bell experiment (CHSH = 2.42) was fed into the modified PBR protocol and the results are presented in figure 4.6.

From figure 4.6, it can be observed that, like the previous test, all graphs show positive linear relationship between $-\log p$ and the number of trials. The difference in this example is that the graph flattens to 0 when the hypothesis CHSH hits 2.5. This implies that there exist some experimental outcomes in the experiment data to have a CHSH value of greater than 2.4. This shows consistency between the true parameter of the state and the analysis result.

Indeed, the $-\log p$-values stays at 0 when the hypothesis CHSH is greater than the CHSH of the experimental data. This is shown in the graph where the test CHSH is greater than 2.4. These results presented in experiment A and B can conclude that the modified PBR protocol has the capability to

inform user about the extent of the Bell violation in the experimental data.

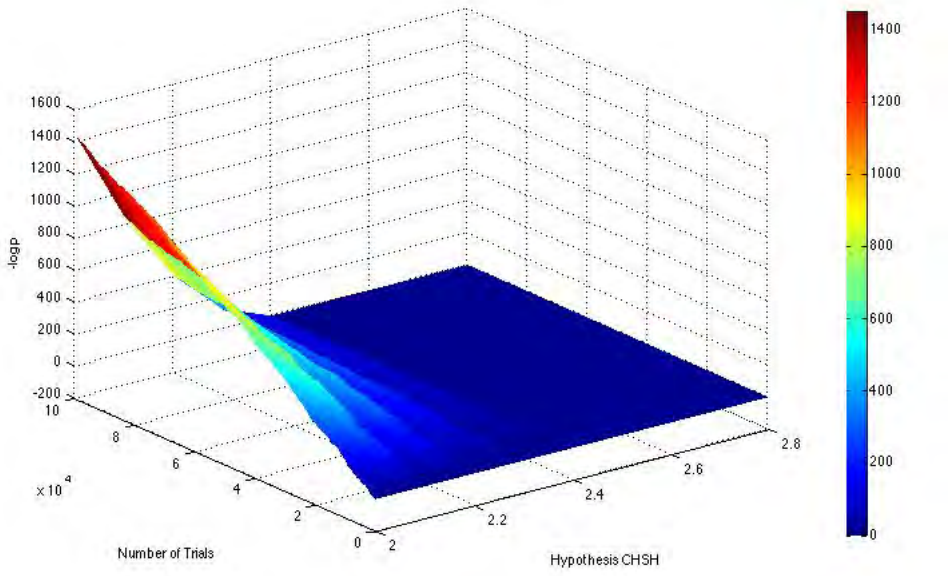

Figure 4.6: The surface plot of $-\log p$-values against number of trials against the hypothesis CHSH. The surface plot is generated by running the experimental data of an IID Bell experiment with a CHSH value of 2.42 on the modified PBR protocol.

Randomness from Experiment A

Now that the validity of the modified protocol has been established, certification of private randomness with modified PBR protocol can proceed.

Recall that the p-value is the maximum probability to obtain the observed or more extreme statistics given that H_0 is true. H_0 will be rejected if the p-value is smaller than the significant value, α. Also, recall that the value of α is arbitrary and is the probability which H_0 is rejected when it should not be.

For the purpose of this discussion, the α is set to be 0.001, which means that for 99.9% of the time, the rejection of H_0 is done correctly. This means that to reject H_0, $-\log p$-value has to be at least 10.0.

The choice of using any hypothesis test is equally valid with the exception of $H_{0,0}$ because the proof of a Bell violation with only prove the presence of private randomness, and is just limited to that. For the purpose of discussion, $H_{0,5}$ will be used to quantify randomness from the experi-

ment outcome. Since the hypothesis CHSH of $H_{0,5}$ is given by 2.5, each experimental outcome with CHSH of 2.5 has a min-entropy of 0.268 bit.

When running an experimental data on the modified PBR protocol with hypothesis CHSH of 2.5, if the $-\log p$-value hits 10.0, it implies that with 99.9% confidence that there exist at least 1 outcome which has CHSH value more than 2.5. This means that with 99.9% confidence that within the outcomes which contributes to the $-\log p$-value of 10.0, there exist at least 0.268 bit of randomness.

The proposed scheme to quantify private randomness for non-IID Bell experiment involves running the modified PBR protocol and reseting the $-\log p$-value to 0 once it exceeds 10. The number of times the $-\log p$-value exceeds 10.0 will be multiplied by 0.268 bit which gives the amount of private randomness present.

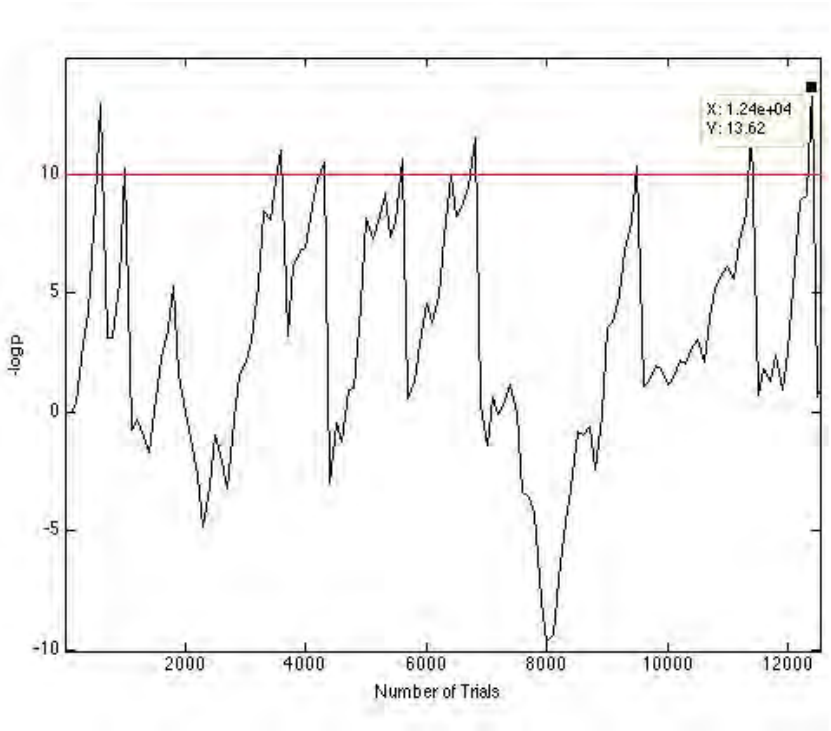

Figure 4.7: Plot of $-\log p$ against the number of trials of experiment A. In this graph, the $-\log p$ resets to 0 when it exceeds 10. The red line visually marks the $-\log p$ of 10.

The results of the proposed scheme on experiment A is shown in figure 4.7. According to the figure, the $-\log p$-value exceeds 10 on 9 occasions after 12,400 trials. This implies that after 12,400 trials, the scheme certifies

at least 2.412 bits of randomness are present with 99.9% confidence.

In fact, there are a total of 81 occasions when $-\log p$-value exceeds 10 in the full run of 100,000 trials. Thus, it is certified that there exist 21.708 bits of private randomness in experiment A with 99.9% confidence. In contrast, if the assumption of IID is valid, it is apparent that there are 100,000 bits of randomness present in the same experimental outcomes.

Chapter 5

Summary and Future Works

5.1 Summary

Bell experiment is able to produce private randomness which can be verified by observing the violation of Bell inequality. Computation of the CHSH value, which is a linear combination of the outcome correlations, is essential to prove a Bell violation.

However, it was shown that the validity of the conventional computation of CHSH value only extends within the realm of IID Bell experiments. Fortunately, it was verified in this paper that the PBR protocol provides a viable alternative to prove the existence of Bell violation in non-IID Bell experiments.

It is clear that the PBR protocol by itself does not give clear instructions on obtaining the estimated probability distribution of the settings and outcomes, $q_x^{(k)}$. A different data half-life value, λ_d, will give rise in a different $q_x^{(k)}$ which may result in totally opposite conclusions given by the protocol. In this way, the choice of $q_x^{(k)}$ remains free and different p-values maybe obtained by different data analysts using the same data. A possible standardisation could be done by taking the minimum p-value over all λ_d to the order of magnitude.

That being said, the sole purpose of the PBR protocol is to provide proof for the presence of Bell violation in experiment and the protocol do not provide users sufficient information to study the extent of Bell violation of a non-IID Bell experiment. This makes quantifying the amount of private randomness in the outcome of Bell experiment impossible using the

PBR protocol. Fortunately, by making the appropriate modifications to the protocol, the user will be able to obtain a lower bound of the amount of private randomness possessed by the experimental outcome of a non-IID Bell experiment.

5.2 Future Works

The quest to quantify private randomness produced by non-IID Bell experiment is far from over. In order to leave no stone unturned, it is important to seek out other feasible methods while working on the current ones.

While working on the "2-days" experiment (Situation A), it is noticed that by re-scrambling the order of the trials using pseudo-random numbers, the p-value remains constant with a small degree of fluctuation. However, the difference is that the graph of $-\log p$ against number of trials for the scrambled data is linear, resembling the IID cases.

This should not be surprising because essentially while scrambling the sequence of the data, the average outcome correlations remains the same. On the other hand, it is highly probable the experimental data from day 1 and day 2 are evenly spread over all data blocks. Hence, it is indistinguishable between the experimental data of an IID Bell experiment with CHSH value of $\frac{2\sqrt{2}}{10} + \frac{18}{10}$ and the scrambled data from "2-days" experiment. Obvious that in both cases, the experiment outcomes contains the same amount of randomness.

Given the above facts, it is reasonable to postulate that by analysing data possessing the same amount of randomness with PBR protocol may give rise to the same p-value. However, concrete studies and additional evidences are required before such claims can be made.

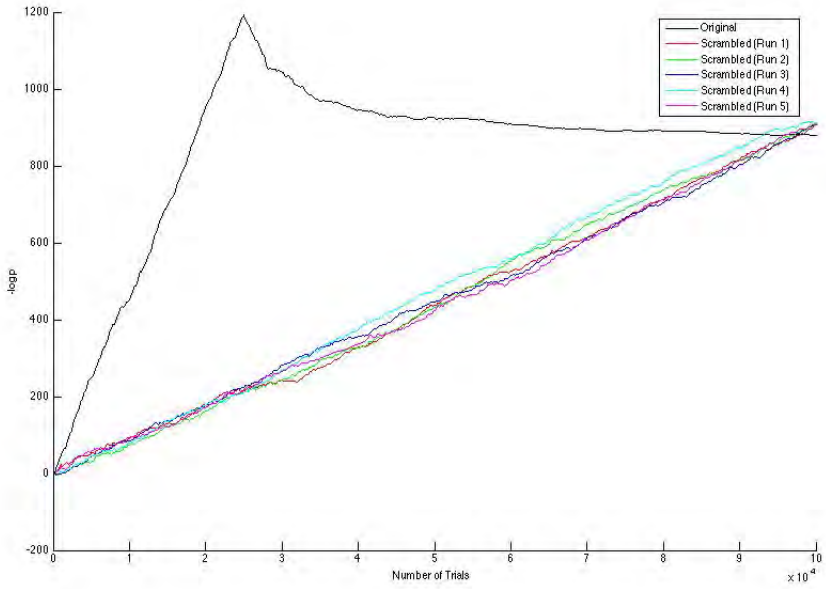

Figure 5.1: Graph of $-\log p$ against the number of trials for "2-days" experiment (Situation A). The graph in black is generated using the original data while the others are generated by re-shuffled data.

Appendix A

List of inequalities bounding "test" polytope

For the purpose of computing the p_{op} in the modified PBR protocol, several inequality constraints are required to be fed into PORTA in order to obtain the extremal points of the expanded polytope. The constraints are as given:

1. $\Pr(a = 1, b = 1) \geq 0$
2. $\Pr(a = -1, b = 1) \geq 0$
3. $\Pr(a = 1, b = -1) \geq 0$
4. $\Pr(a = -1, b = -1) \geq 0$
5. $\Pr(b = 1) - \Pr(a = -1, b = 1) \geq 0$
6. $\Pr(b = 1) - \Pr(a = 1, b = 1) \geq 0$
7. $\Pr(b = -1) - \Pr(a = -1, b = -1) \geq 0$
8. $\Pr(b = -1) - \Pr(a = 1, b = -1) \geq 0$
9. $\Pr(a = 1) - \Pr(a = 1, b = -1) \geq 0$
10. $\Pr(a = 1) - \Pr(a = 1, b = 1) \geq 0$
11. $\Pr(a = -1) - \Pr(a = -1, b = -1) \geq 0$
12. $\Pr(a = -1) - \Pr(a = -1, b = 1) \geq 0$

13. $-\Pr(a=-1)-\Pr(b=-1)+\Pr(a=-1,b=-1)\geq -1$

14. $-\Pr(a=-1)-\Pr(b=1)+\Pr(a=-1,b=1)\geq -1$

15. $-\Pr(a=1)-\Pr(b=-1)+\Pr(a=1,b=-1)\geq -1$

16. $-\Pr(a=1)-\Pr(b=1)+\Pr(a=1,b=1)\geq -1$

17. $\Pr(a=1)+\Pr(a=-1,b=-1)-\Pr(a=1,b=-1)+\Pr(b=1)-\Pr(a=-1,b=1)-\Pr(a=1,b=1)\geq -\frac{\text{CHSH}-2}{4}$

18. $\Pr(a=1)+\Pr(b=-1)-\Pr(a=-1,b=-1)-\Pr(a=1,b=-1)+\Pr(a=-1,b=1)-\Pr(a=1,b=1)\geq -\frac{\text{CHSH}-2}{4}$

19. $\Pr(a=-1)-\Pr(a=-1,b=-1)+\Pr(a=1,b=-1)+\Pr(b=1)-\Pr(a=-1,b=1)-\Pr(a=1,b=1)\geq -\frac{\text{CHSH}-2}{4}$

20. $\Pr(a=-1)+\Pr(b=-1)-\Pr(a=-1,b=-1)-\Pr(a=1,b=-1)-\Pr(a=-1,b=1)+\Pr(a=1,b=1)\geq -\frac{\text{CHSH}-2}{4}$

21. $-\Pr(a=-1)-\Pr(b=-1)+\Pr(a=-1,b=-1)+\Pr(a=1,b=-1)+\Pr(a=-1,b=1)-\Pr(a=1,b=1)\geq -\frac{\text{CHSH}+2}{4}$

22. $-\Pr(a=-1)+\Pr(a=-1,b=-1)-\Pr(a=1,b=-1)-\Pr(b=1)+\Pr(a=-1,b=1)+\Pr(a=1,b=1)\geq -\frac{\text{CHSH}+2}{4}$

23. $-\Pr(a=1)-\Pr(b=-1)+\Pr(a=-1,b=-1)+\Pr(a=1,b=-1)-\Pr(a=-1,b=1)+\Pr(a=1,b=1)\geq -\frac{\text{CHSH}+2}{4}$

24. $-\Pr(a=1)-\Pr(a=-1,b=-1)+\Pr(a=1,b=-1)-\Pr(b=1)+\Pr(a=-1,b=1)+\Pr(a=1,b=1)\geq -\frac{\text{CHSH}+2}{4}$

Constraints 1 to 4 are the positivity constraints of probabilities which can be rewritten as $\Pr(ab)\geq 0$. Constraints 5 to 12 demand that the conditional probabilities $\Pr(a|b)$ and $\Pr(b|a)$ has to be less than or equals to 1. Constraints 13 to 16 demand that the probabilities of outcomes a and/or b occurring has to be less than or equals to 1. (ie. $\Pr(a\cup b)\leq 1$) Constraints 17 to 24 are the CH inequalities which are equivalent to the CHSH inequality but it deals with probabilities rather than correlations.

Bibliography

[1] Antonio Acin, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters*, 98(23):230501, 2007.

[2] Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu. Quantum nonlocality, bell inequalities, and the memory loophole. *Physical Review A*, 66(4):042111, 2002.

[3] John Stuart Bell. On the einstein-podolsky-rosen paradox. physics 1, 195–200 (1964). reprinted in js bell, speakable and unspeakable in quantum mechanics, 1987.

[4] Thomas Christof, Andreas Löbel, and M Stoer. Porta-polyhedron representation transformation algorithm. *Software package, available for download at http://www. zib. de/Optimization/Software/Porta*, 1997.

[5] Boris S Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[6] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23:880–884, 1969.

[7] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

[8] Richard D Gill. Time, finite statistics, and bell's fifth position. *arXiv preprint quant-ph/0301059*, 2003.

[9] Richard D Gill. Statistics, causality and bell's theorem. *arXiv preprint arXiv:1207.5103*, 2012.

[10] Pawel Kurzynski, Marcin Markiewicz, and Dagomir Kaszlikowski. On compression of non-classically correlated bit strings. *arXiv preprint arXiv:1310.5644v1*, 2013.

[11] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzimitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al. Random numbers certified by bells theorem. *Nature*, 464(7291):1021–1024, 2010.

[12] Jiangwei Shang, Hui Khoon Ng, Arun Sehrawat, Xikun Li, and Berthold-Georg Englert. Optimal error regions for quantum state estimation. *New Journal of Physics*, 15(12):123026, 2013.

[13] Wim Van Dam, Richard D Gill, and Peter D Grunwald. The statistical strength of nonlocality proofs. *Information Theory, IEEE Transactions on*, 51(8):2812–2835, 2005.

[14] Y Vardi and D Lee. From image deblurring to optimal investments: Maximum likelihood solutions for positive linear inverse problems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 569–612, 1993.

[15] Yanbao Zhang, Scott Glancy, and Emanuel Knill. Asymptotically optimal data analysis for rejecting local realism. *Physical Review A*, 84(6):062118, 2011.