QUANTUM INFORMATION THEORY

The CHSH Inequality

Stuti Shah
Kapitza Society, Spring 2022, Prof. S.G. Rajeev
University of Rochester

April 27, 2022

1 Introduction

This paper is a compilation of the content I covered in two of the lectures I gave for Kapitza meetings this semester. We continued studying quantum information from last semester and in this paper we begin by looking at the CHSH inequality, followed by the CHSH game, and finally a different topic yet one with remote similarities with the previously mentioned topics, a quantum analog of Shannon's noiseless coding theorem.

2 The CHSH Inequality

Before we look at the CHSH inequality, let us familiarize ourselves with three assumptions [1] that will hold true throughout our discussion of classical correlations:

Assumption 1:

The quantities we choose to measure already have certain values prior to any measurement.

This means that when we make a measurement, we are only attempting to determine the value and not assign any value - once we know what the outcome is, after the measurement is made - to those quantities. Note that it is not necessary that those pre-contained values are known to us before we measure the quantities, and therefore we might require a formalism developed using probabilistic arguments to describe the experiment.

Assumption 2:

The properties of one subsystem should not immediately depend on the events taking place on a very distant subsystem.

Suppose we have a system consisting of a source that emits particles, and two detectors (far apart from the source and each other) that measure the particles. The statement says that the measurements made on one detector should be independent of those made on the other. Say, two detectors A and B each perform ± 1 valued measurements, and the probability that detector A measures outcome $a \in \{\pm 1\}$ and detector B measures outcome $b \in \{\pm 1\}$ is $p(a, b|x, y)$, then,

$$\langle A_x B_y \rangle \equiv \sum_{a,b \in \{\pm 1\}} ab \, p(a, b|x, y)$$

where x and y in $\mathbb{N}$ are measurement devices.

Assumption 3:

Let $A_x, B_y : \Omega \to \{\pm 1\}$ be random variables and P a probability measure. We have the expectation of the product to be: $\langle A_x B_y \rangle = \int_\Omega A_x(\omega) B_y(\omega) dP(\omega)$.

[1] Lachenmaier M., Michel M., Classical vs. Quantum Correlations, 2016

ω here is the unknown or 'hidden' variable from assumption 1, and A_x and B_y do not depend on y and x respectively.

The Causer-Horne-Shimony-Holt (CHSH) Inequality

Statement: The description of ± 1 valued measurements made by the devices A_1, A_2, B_1, B_2, satisfies

$$|\langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_2\rangle - \langle A_2 B_2\rangle| \leq 2 \tag{1}$$

Proof [2] : From assumption 3, we have,

$$\langle A_1 B_1\rangle = \int_\Omega A_1(\omega)B_1(\omega)dP(\omega)$$

$$\langle A_1 B_2\rangle = \int_\Omega A_1(\omega)B_2(\omega)dP(\omega)$$

$$\langle A_2 B_1\rangle = \int_\Omega A_2(\omega)B_1(\omega)dP(\omega)$$

$$\langle A_2 B_2\rangle = \int_\Omega A_2(\omega)B_2(\omega)dP(\omega)$$

So the left hand side of (1) becomes,

$$\begin{aligned} LHS &= \left|\int_\Omega A_1(\omega)B_1(\omega)dP(\omega) + \int_\Omega A_1(\omega)B_2(\omega)dP(\omega) + \int_\Omega A_2(\omega)B_1(\omega)dP(\omega) - \int_\Omega A_2(\omega)B_2(\omega)dP(\omega)\right| \\ &= \left|\int_\Omega [A_1(\omega)B_1(\omega) + A_1(\omega)B_2(\omega) + A_2(\omega)B_1(\omega) - A_2(\omega)B_2(\omega)]dP(\omega)\right| \\ &= \left|\int_\Omega [A_1(\omega)\{B_1(\omega) + B_2(\omega)\} + A_2(\omega)\{B_1(\omega) - B_2(\omega)\}]dP(\omega)\right| \end{aligned}$$

$A_1(\omega)$ and $A_2(\omega)$ are in $\{\pm 1\}$.

For a particular $\omega \in \Omega$, we can have two conditions:

1. $B_1(\omega) = B_2(\omega)$

2. $B_1(\omega) \neq B_2(\omega)$

For case 1., $B_1(\omega) - B_2(\omega) = 0$ and $B_1(\omega) + B_2(\omega) \in \{\pm 2\}$.

For case 2., since either 1 or -1 is the only value possible, $B_1(\omega) = -B_2(\omega)$; that is, $B_1(\omega) + B_2(\omega) = 0$ and $B_1(\omega) - B_2(\omega) \in \{\pm 2\}$. Hence, the desired result is acheived.

3 The CHSH Game

"The whole point of the CHSH game [and Bell's inequality] is to prove that quantum entanglement exists! The CHSH game shows us how the physical systems of the universe are more aligned to, and better represented by, quantum physics rather than classical physics." (Eamonn Darcy)

The CHSH game is a two-player game consisting of players Alice and Bob who each receive a bit $x \in \{0, 1\}$ and $y \in \{0, 1\}$ as an input (or a "question") from the referee Charlie, respectively. Both the players must send an output in response to the question to Charlie, without communicating in any manner with each other (they know beforehand that both their inputs are chosen at random from $\{0, 1\}$ i.e. all of the 4 possible input pairs (0,0), (0,1), (1,0), (1,1) are equally likely). Say, Alice's answer is a and that of Bob is b. The task is for Alice and Bob to give matching outputs for each of the questions asked (i.e. $a = b$) except when the question is (1,1) (where their outputs must be $a \neq b$). That is, after receiving both the answers, Charlie determines whether the players won or lost the game, which means it is impossible for one to win and the other to lose.

With the same format of the game just described, we can write the condition on the outputs that Alice and Bob give, mathematically, as follows:

$$x \cdot y = a \oplus b = a + b \bmod 2$$

[2] Lachenmaier M., Michel M., Classical vs. Quantum Correlations, 2016

3.1 Classical strategy

A little bit of thought should convince us that in any classical scenario, Alice and Bob have a $\leq 75\%$ chance of winning.

Let us look at a proof using the CHSH inequality[3]:

Suppose that Alice and Bob return values in $\{\pm 1\}$ instead of $\{0, 1\}$. For questions x and y in $\{0, 1\}$, let's say Alice and Bob answer r_A and r_B in $\{\pm 1\}$ and that they returned a and b in $\{0, 1\}$ originally. Then, we have:

(x, y)	$a \oplus b$	$r_A r_B$
(0,0)	0	1
(0,1)	0	1
(1,0)	0	1
(1,1)	1	-1

Consider measuring devices A_0 (for $x = 0$) and A_1 (for $x = 1$) for Alice and B_0 (for $y = 0$) and B_1 (for $y = 1$) for Bob. For $x, y \in \{0, 1\}$, from assumption 2,

$$\langle A_x B_y \rangle = \sum_{r_A, r_B \in \{\pm 1\}; r_A = r_B} p(r_A, r_B | x, y) - \sum_{r_A, r_B \in \{\pm 1\}; r_A \neq r_B} p(r_A, r_B | x, y)$$

The above equation gives us difference in the probability of winning and losing for inputs equal to and not equal to (1,1).

$$\begin{aligned} P_{\text{win}} - P_{\text{lose}} &= \frac{1}{4}[\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle] \leq \frac{1}{2} \\ P_{\text{win}} - (1 - P_{\text{win}}) &\leq \frac{1}{2} \\ P_{\text{win}} &\leq \frac{\frac{1}{2} + 1}{2} = \frac{3}{4} = 75\% \end{aligned}$$

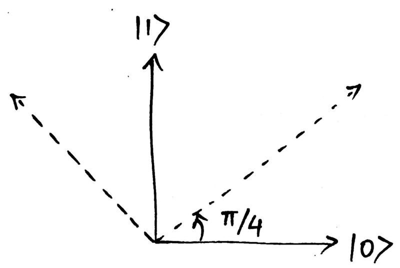
3.2 Quantum strategy

What happens if Alice and Bob were to share a Bell pair (i.e. if Alice had one qubit of the Bell state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and Bob had the other)? Are their chances of winning better? Quantum mechanics predicts that Alice and Bob will succeed with 85% instead if they use the following strategy:

For $\theta \in [0, 2\pi)$, define [4]

$$|\phi_0(\theta)\rangle \equiv \cos\theta |0\rangle + \sin\theta |1\rangle$$
$$|\phi_1(\theta)\rangle \equiv -\sin\theta |0\rangle + \cos\theta |1\rangle$$

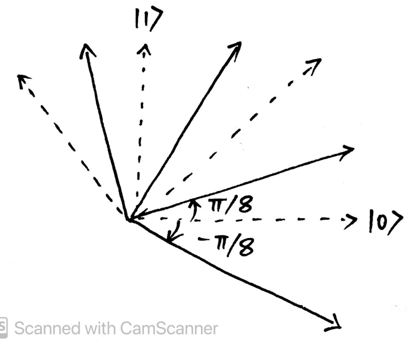
Here's what Alice does: Based on the input given to her, she measures her qubit in one of two bases,



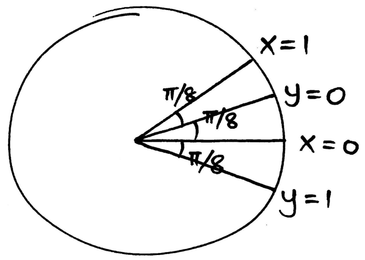
[3]Lachenmaier M., Michel M., Classical vs. Quantum Correlations, 2016
[4]Lachenmaier M., Michel M., Classical vs. Quantum Correlations, 2016

If her input is 0, she measures her qubit in the $\{|\phi_0(0)\rangle, |\phi_1(0)\rangle\}$ basis.
If her input is 1, she measures her qubit in the $\{|\phi_0(\frac{\pi}{4})\rangle, |\phi_1(\frac{\pi}{4})\rangle\}$ basis.

Here's what Bob does: Based on the input given to him, he measures his qubit in one of two bases,



If his input is 0, he measures his qubit in the $\{|\phi_0(\frac{\pi}{8})\rangle, |\phi_1(\frac{\pi}{8})\rangle\}$ basis.
If his input is 1, he measures his qubit in the $\{|\phi_0(-\frac{\pi}{8})\rangle, |\phi_1(-\frac{\pi}{8})\rangle\}$ basis.



Recall the rotational invariance of bell state: if one measures a qubit in a certain basis and measures the other qubit in a basis that is rotated by θ with respect to the original basis, then the probability of getting the same outcome, $P_{\text{same}} = \cos^2\theta$.

x	y	P_{same}	$P_{\text{different}}$
0	0	$\cos^2\frac{\pi}{8}$	-
0	1	$\cos^2\frac{\pi}{8}$	-
1	0	$\cos^2\frac{\pi}{8}$	-
1	1	$\cos^2\frac{3\pi}{8}$	$1 - P_{\text{same}} = 1 - \cos^2\frac{3\pi}{8} = \sin^2\frac{3\pi}{8} = \cos^2\frac{\pi}{8}$

Thus, the probability of getting the correct answer (i.e. meeting the rules of the game) is $\cos^2\frac{\pi}{8} = 0.85 = 85\%$ for all the 4 cases. Thus, if Alice and Bob incorporate quantum strategies as opposed to classical strategies to play this game, they have a better chance at winning.[5]

4 Quantum analog of Shannon's noiseless coding theorem [6]

There are two things [7] that Shannon's noiseless coding theorem tells us:

1. It is impossible to compress data ∍ the code rate, i.e. $\frac{\text{average no. of bits}}{\text{symbol}}$ < Shannon entropy of the source, without any certainty that no information will be lost in the process.

2. It is possible to get the code rate arbitrarily close to Shannon entropy with negligible probability of loss.

[5]Lecture 4 3 chsh inequality - youtube, (n.d.), Retrieved, 2022
[6]Preskill, J., Lecture Notes for Physics 229: Quantum Information and Computation, 1998
[7]Wikimedia Foundation, March 30, 2022, Shannon's source coding theorem, Wikipedia, Retrieved 2022

In this section, we look at quantum analog of Shannon's noiseless coding theorem. The content in this section was the topic of my second lecture for Kapitza this semester. A reason to include this topic in this paper is to illustrate that similar numbers appear in the discussion of different topics.

Let's say we have a message comprising of n letters that we want to deliver. There exists no bias and each n has an equal probability of being chosen, from an ensemble of pure states described by

$$\{|\psi_x\rangle, p_x\}$$

Note that the $|\psi_x\rangle$'s do not have to be necessarily orthogonal (for example, as I described in my earlier paper on density matrices, the polarization state of a photon). Based on arguments drawn from my previous paper, each letter has a density matrix

$$\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x| \tag{2}$$

to fully describe it. Since our message consists of n letters, the density matrix for the entire message is just

$$\rho^n = \rho \otimes \rho \otimes ... \otimes \rho$$

(n times). Our task here, says Preskill, is to compress this message (say for conserving space on a quantum computer hard disk when statistical properties - ρ - of the recorded data are known) to a smaller Hilbert space $\mathcal{H}$ without compromising the *fidelity* of the message. An optimal compression would be one with arbitrarily good fidelity as $n \to \infty$. Such a compression was found by Ben Schumacher:

$$\log(\dim \mathcal{H}) = nS(\rho)$$

where the Von Neumann entropy $S(\rho)$ here is the number of qubits of information carried per letter of the message.

Let us consider an example to understand this better. Suppose we "alias" 'letters' with 'single qubits' and they're chosen from the following ensemble:

$$|\uparrow_z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; p_z = \frac{1}{2} \tag{3}$$

$$|\uparrow_x\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}; p_x = \frac{1}{2} \tag{4}$$

From (2) we have,

$$\rho = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

The eigenstates of ρ are qubits oriented up and down along the axis $\hat{n} = \frac{1}{\sqrt{2}}(\hat{x} + \hat{z})$; we define

$$|0'\rangle \equiv |\uparrow_{\hat{n}}\rangle = \begin{pmatrix} \cos\frac{\pi}{8} \\ \sin\frac{\pi}{8} \end{pmatrix} \tag{5}$$

$$|1'\rangle \equiv |\downarrow_{\hat{n}}\rangle = \begin{pmatrix} \sin\frac{\pi}{8} \\ -\cos\frac{\pi}{8} \end{pmatrix} \tag{6}$$

Recall that we saw similar notation when we studied the Bloch sphere last semester! The eigenvalues, found using the $\det(\rho - \lambda\mathbb{I})$ computation, are

$$\lambda(0') = \frac{2+\sqrt{2}}{4} = 0.8535 = \cos^2\frac{\pi}{8}$$

$$\lambda(1') = \frac{2-\sqrt{2}}{4} = 0.1465 = \sin^2\frac{\pi}{8}$$

Before moving on, notice that we saw the 0.85 value in the CHSH game above!

Two things to observe here; first,

$$\lambda(0') + \lambda(1') = \cos^2\frac{\pi}{8} + \sin^2\frac{\pi}{8} = 1$$

$$\begin{aligned}\lambda(0')\cdot\lambda(1') &= \cos^2\frac{\pi}{8}\cdot\sin^2\frac{\pi}{8}\\ &= \frac{1}{4}\left(2\sin\frac{\pi}{8}\cos\frac{\pi}{8}\right)^2\\ &= \frac{1}{4}\left(\sin\frac{\pi}{4}\right)^2\\ &= \frac{1}{8}\\ &= \det\rho\end{aligned}$$

and second, from (3), (4), (5), and (6):

$$|\langle 0'|\uparrow_z\rangle|^2 = |\langle 0'|\uparrow_x\rangle|^2 = \cos^2\frac{\pi}{8} = 0.8535$$

$$|\langle 1'|\uparrow_z\rangle|^2 = |\langle 1'|\uparrow_x\rangle|^2 = \sin^2\frac{\pi}{8} = 0.1465$$

Equalities (5) and (6) tell us that the $|0'\rangle$ state has equal and large overlap with both the signal states and the $|1'\rangle$ state has equal and small overlap with both the signal states, which suggests that if the input signal state is unknown to us, the best guess we can make is $|\psi_x\rangle = |0'\rangle$ and the $|0'\rangle$ state has the maximal fidelity,

$$F = \frac{1}{2}|\langle\uparrow_z|\psi\rangle|^2 + \frac{1}{2}|\langle\uparrow_x|\psi\rangle|^2 = \frac{1}{2}(0.8535) + \frac{1}{2}(0.8535) = 0.8535$$

Let us talk about a more specific example now that we've established some background. Say that Alice has a message that consists of three letters that she wants to send to Bob. However, she can only afford to send two of the three letters since quantum communication is expensive. The goal is for Bob to construct Alice's message (state) with maximum correctness, or *fidelity.* Since Alice discloses two of the letters in her message to Bob, the two letters have $F = 1$. From our previous understanding, we can figure that Bob guesses $|0'\rangle$ ($F = 0.8535$) in order to try and construct Alice's message with maximum fidelity. The total fidelity is therefore $F = 0.8535$. This seems pretty good, although, there is a better strategy for Bob to guess Alice's message with even more correctness.

Let us make our methods explicit:
What we did: We decomposed the $\mathcal{H}$ of one qubit into a "likely" subspace spanned by $|0'\rangle$ and an "unlikely" subspace spanned by $|1'\rangle$.
What we can do now: We can decompose the $\mathcal{H}$ of three qubits (since Alice's message contains 3 'letters') into "likely" and "unlikely" subspaces. Let the input signal state $|\psi\rangle$ be expressed as

$$|\psi\rangle = |\psi_1\rangle|\psi_2\rangle|\psi_3\rangle$$

with $|\psi_1\rangle$, $|\psi_2\rangle$, $|\psi_3\rangle$ in either $|\uparrow_z\rangle$ or $|\uparrow_x\rangle$ state. We therefore have, with 2 axes and 3 qubits, $2^3 = 8$ possibilities:

$$|\langle 0'0'0'|\psi\rangle|^2 = \left(\cos^2\frac{\pi}{8}\right)^3 = \cos^6\frac{\pi}{8} = 0.6219 \tag{7}$$

$$|\langle 0'0'1'|\psi\rangle|^2 = |\langle 0'1'0'|\psi\rangle|^2 = |\langle 1'0'0'|\psi\rangle|^2 = \cos^4\frac{\pi}{8}\sin^2\frac{\pi}{8} = 0.1067 \tag{8}$$

$$|\langle 0'1'1'|\psi\rangle|^2 = |\langle 1'0'1'|\psi\rangle|^2 = |\langle 1'1'0'|\psi\rangle|^2 = \cos^2\frac{\pi}{8}\sin^4\frac{\pi}{8} = 0.0183 \tag{9}$$

$$|\langle 1'1'1'|\psi\rangle|^2 = \left(\sin^2\frac{\pi}{8}\right)^3 = \sin^6\frac{\pi}{8} = 0.0031 \tag{10}$$

From this, we can decompose the space into the "likely" subspace Λ spanned by $\{|0'0'0'\rangle, |0'0'1'\rangle, |0'1'0'\rangle, |1'0'0'\rangle\}$ and the "unlikely" subspace $\Lambda^\perp$ spanned by $\{|0'1'1'\rangle, |1'0'1'\rangle, |1'1'0'\rangle, |1'1'1'\rangle\}$.

In order to make a ("fuzzy") measurement, Alice applies a unitary transformation U that rotates the high probability states $|0'0'0\rangle$, $|0'0'1'\rangle$, $|0'1'0'\rangle$, and $|1'0'0'\rangle$ to $|\rangle|\rangle|0\rangle$ and the four low probability states $|0'1'1'\rangle$, $|1'0'1'\rangle$, $|1'1'0'\rangle$, and $|1'1'1'\rangle$ to $|\rangle|\rangle|1\rangle$. She then measures the third qubit. Two things could happen here:

Case 1: She measures the third qubit to be $|0\rangle$, which would mean that her input state has been projected to Λ. From (7) and (8),

$$P_\Lambda = 0.6219 + 3(0.1067) = 0.9419$$

Case 2: She measures the third qubit to be $|1\rangle$, which would mean that her input state has been projected to $\Lambda^{\perp}$. From (9) and (10),

$$P_{\Lambda^{\perp}} = 3(0.0183) + 0.0031 = 0.0581$$

For case 1, Alice sends the remaining (unmeasured, compressed) two qubits $|\psi_{\text{comp}}\rangle$ to Bob, who decompresses it to obtain:

$$|\psi'\rangle = U^{-1}(|\psi_{\text{comp}}\rangle|0\rangle)$$

For case 2, Alice sends a state that Bob would decompress to get the most likely state. Thus:

$$|\psi'\rangle = U^{-1}(|\psi_{\text{comp}}\rangle|0\rangle) = |0'0'0'\rangle$$

Through this process, Bob finds his state ρ' to be:

$$\rho' = |\psi'\rangle\langle\psi'| = E|\psi\rangle\langle\psi|E + |0'0'0'\rangle\langle\psi|1 - E|\psi\rangle\langle 0'0'0'|$$

where E is the projection onto Λ. Hence,

$$\begin{aligned} F &= \langle\psi|\rho'|\psi\rangle \\ &= (\langle\psi|E|\psi\rangle)^2 + (\langle\psi|1 - E|\psi\rangle)(\langle\psi|0'0'0'\rangle)^2 \\ &= (0.9419)^2 + (0.0581)(0.6219) \\ &= 0.9234 \end{aligned}$$

Clearly, the fidelity in this case (0.9234) is better than before (0.8535). It should only be true that as the number of letters in a message increases, the fidelity of the compressed message should increase too.

$$S(\rho) = H \cos^2 \frac{\pi}{8} \approx 0.6009$$

for one-qubit ensemble, which suggests that we can compress the incoming message by a factor of 0.6009 and still retain good, reliable accuracy.

5 References

Preskill, J., Lecture Notes for Physics 229: Quantum Information and Computation, 1998

Lachenmaier M., Michel M., Classical vs. Quantum Correlations, 2016

Lecture 4 3 chsh inequality - youtube, (n.d.), Retrieved, 2022

Wikimedia Foundation, March 30, 2022, Shannon's source coding theorem, Wikipedia, Retrieved 2022