Mutual Information: A way to quantify correlations

Información Mutua: Una forma de cuantificar las correlaciones

Marcelo Tisoc*[1][ORCID], Jhosep Victorino Beltrán[1]

[1]Universidad Nacional de Ingenieria, Facultad de Ciencias, Lima, Peru.

Within the framework of Information Theory, the existence of correlations between two random variables means that we can obtain information about one of them, just by measuring or observing the other random variable. In certain cases, this kind of relationship allows obtaining information about a variable even when the other is separated by a very large distance, that is, the process of obtaining information is non-local, an example (if not the only) is the quantum entanglement. These features of correlations make it interesting and important to study, classify and quantify them. The correlations are classified into classical correlations and quantum correlations, in addition they are quantified through the mutual information. Here we will present a natural way to define classical mutual information and then we will generalize it to the quantum case. Furthermore, every term in the definitions of mutual information will be interpreted using the concepts of classical and quantum entropy.
Keywords: Information entropy, Correlations, Mutual information.

Dentro del marco de la Teoría de la información, que exista correlación entre dos variables aleatorias significa que podemos obtener información de una de ellas, tan solo midiendo u observando a la otra variable. En ciertos casos, este tipo de relación permite obtener información de una variable aún cuando la otra esté separada una distancia muy grande, es decir que el proceso de obtención de información puede ser no local, un ejemplo (si no el único) es el entrelazamiento cuántico. Este tipo de características hace que sea interesante e importante estudiar, clasificar y cuantificar las correlaciones. Estas se clasifican en correlaciones clásicas y correlaciones cuánticas, además se cuantifican a través de la información mutua, definida tanto como para sistemas clásicos como para sistemas cuánticos. Presentaremos un camino natural a la definición de la información mutua clásica para luego generalizarla al caso cuántico. Además, cada término en las definiciones de información mutua será interpretado usando los conceptos de entropía clásica y cuántica.
Palabras clave: Entropía de la información, Correlaciones, Información mutua.

1. Introduction

The concept of information is too broad to be captured completely by a single definition. Furthermore, questions about what information is, can easily lead us down the dark path of metaphysics. Andrade says [1]: "Information in its semantic connotation is creations of meanings, which means that it becomes evident when it has been received and has caused some kind of modification in the receptor. In a pragmatic way, it can be said that it is every difference that makes a difference". Hence, since only the differences allow us to perceive the information, we will not be able to quantify the information directly but rather through its variation, it means, through the gain or lack of information. Therefore a measure of information should satisfy this intuitive notions about information [2]. Shannon defined a quantity called Shannon entropy [3–5] which has many properties (see the general properties of entropy in ref. [6]) that agree with a consistent measure of information. For any probability distribution, Shannon entropy also called classical entropy, measures the lack of information to determine the outcome of that probability distribution.

We can use these ideas, for example, to measure the lack of information to determine a message consisting of letters from a certain vocabulary. In this case, each letter of the message is independent from the other. However, if we consider that the message was sent in a certain language (Shannon did that analysis for English, see Ref. [7]), hence we would be introducing constraints on the letters of the message. Let us see an example to be clear, suppose that we received a message consisting of letters from a vocabulary and we know that the first letter is "p", knowing this will not give us information about the next letter. In other words, our lack of information about the second letter will not vary after know the first letter. However, if we know that the message was sent in English and the first letter of the message is "p", if we know English then we will know that the next letter must be a vocal, i.e., "a, e, i, o, u" or the letters "r", "l" or "s" and not other. This means that knowing the first letter (letter "p") allow us to reduced the uncertainty about the second letter. To visualize better the idea let us see the next extreme case:

* Correspondence email address: marcelo.tisoc.m@uni.pe

In English, if the first letter is "q", then, we know with certainty that the next letter must be "u". Summarizing, in certain languages, the first letter gives us information about the second letter. It means that this two letters are correlated. Hence whether one random variable contains information about other, then both random variables are correlated [8].

The natural question is, how to quantify correlations? The last example shows us that correlations are intimately linked to entropy of information and hence, a consistent measure of correlations must be defined in function of entropy. Mutual Information was proposed to quantify correlations between two random variables [2]. Since in physics we have two kind of systems, i.e., classical systems and quantum systems, hence in this article we will work also with two kind of correlations, classical correlations and quantum correlations [9]. Due to this dichotomy, it is necessary to define the mutual information not only for classical systems but also for quantum systems. For didactic reasons, we will first study Classical Entropy in Section 2.1 and then we will generalize it to the quantum level in Section 2.2.

Finally in Section 3.1 we define two equivalent definitions for classical mutual information and in the next Section 3.2, we generalize those definitions to the quantum case and we will see that they are no longer equivalent.

2. Entropy in Information Theory

2.1. Classical entropy

In order to quantify the information contained in a classical random variable, first we have to answer the following question: Which function should we use to quantify the information contained in an event?

Suppose that we want to quantify the information contained in the event m, which happens with probability p_m. It is worth to note that something that happens frequently (it means that its probability of happening is high) will not be interesting. For example [10], dawn is an usual event, and even though sunrise is beautiful and tells us that the day is beginning, it does not bring us new information. On the other hand if it is 12 noon and suddenly the night begins, someone will be surprised until they noticed that a solar eclipse is happening, this event, that does not occur frequently (it means that its probability of happening is low) is telling us a lot of information. In short, a highly probable event does not bring us too much information while an unlikely event bring us a lot of information, this is because it causes us more surprise. This reasoning tells us that the information H_m contained in a single event m is monotonically decreasing with respect to the probability p_m, i.e., for two events m and n, if $p_m > p_n$ then $H_m < H_n$.

Now that we have the relationship between the information H_m and the probability p_m of a single event, in order to construct a function that quantifies information we have to give a minimum requirements that it must to satisfy. Let's see:

For two independent events m and n, its informations I_m and I_n must be additive, i.e., $I_{mn} = I_m + I_n$. If we consider the events as a random variables, from probability theory [11], the joint probability of two independent random variables is the product of individual probabilities, i.e., $p_{mn} = p_m p_n$. The only function that satisfies this condition is the logarithm function. Then the information of a single event, also called self-entropy, is defined as follows.

Definition 2.1 *The self-entropy of a single event m with probability p_m is*

$$H_m := \log(1/p_m) = -\log(p_m). \tag{1}$$

The self-entropy quantify the surprise or information contained in a single event.

However, as we are interested in a set of events, we will consider a random variable

$$X = \{x_m | m = 1, 2, \ldots, M\}. \tag{2}$$

The law probability, i.e., the probability p_m of occurrence of event x_m is

$$p_m = P(X = x_m), \tag{3}$$

where the probabilities $\{p_m\}$ satisfy the following conditions

$$0 \le p_m \le 1 \quad \text{and} \quad \sum_i^M p_m = 1. \tag{4}$$

Then, the classical entropy of the random variable X is the weighted average of the self-entropies of each single event, it is

$$H(X) = H(p_1, \ldots, p_m, \ldots, p_M) := \sum_m^M p_m H_m. \tag{5}$$

Finally, using equation (1), we get the well known relation of classical entropy, also known as Shannon entropy. This is because it was first introduced by Shannon [5].

Definition 2.2 *The classical entropy $H(X)$ of one discrete random variable X with sample space $\mathcal{X}$ and probability law $P(X = x) = p_x$ is:*

$$H(X) = -\sum_{x \in \mathcal{X}} p_x \log_2(p_x). \tag{6}$$

The Shannon entropy $H(X)$ quantifies the uncertainty or lack of information to determine which event x from the set of events $\mathcal{X}$ of the random variable X will occur.

To have a better insight of Shannon entropy, we will consider the experiment of tossing a coin. This experiment have two possible results, we could obtain heads or tails. We use the random variable X to assign real numbers to each result, for example the number 1 for heads and number -1 for tails. Therefore, we have a probability law who assign the probability $1/2$ to value 1 (heads) and probability $1/2$ to value -1 (tails). We can see a scheme of this experiment in Figure 1a. Here the classical entropy of the random variable X is $H(X) = 1$ and it is quantifying the lack of information about which result we will get after tossing a coin.

But what if we have two random variables? Given two random variables, X with the set of possible outcomes $\{x_i\}$ and Y with the set of possible outcomes $\{y_i\}$, we define the joint probability distribution:

$$P(X = x, Y = y) \equiv p_{xy}. \qquad (7)$$

With corresponding marginal probability distributions

$$P(X = x) = p_x = \sum_y p_{xy},$$
$$P(Y = y) = p_y = \sum_x p_{xy}. \qquad (8)$$

The degree of correlation between the random variable X and Y is encoded in their joint probability distribution and best quantified within the framework of Shannon entropy [9]. The lack of information about the mixture of two random variables is quantified through the joint entropy defined as follows.

Definition 2.3 *The Joint Entropy of two random variables X and Y is:*

$$H(A, B) = -\sum_{xy} p_{xy} \log p_{xy}, \qquad (9)$$

where p_{xy} is the joint probability.

2.1.1. Classical conditional entropy

Now, from classical entropy we will derive an important quantity to study correlations. Its name is *conditional entropy* and to make its definition intuitive we will perform the following thought experiment.

Let's suppose that Alice wants to communicate with Bob and she sends a message to him. This is equivalent to say that Alice sends a random variable X to Bob, the possible outcomes are $\{x\}_{x \in \mathcal{X}}$ with probability law $P(X) = p_x$ where $\mathcal{X}$ is the support[1] of random variable X. She sends the message X through a noisy telephone and therefore what Bob will receive will be the random variable Y whose possible outcomes are $\{y\}_{y \in \mathcal{Y}}$ with probability law $P(Y) = p_y$, where $\mathcal{Y}$ is the support of Y.

---

[1] The support of a random variable is the subset of elements which are not mapped to zero.

The natural question that arises is: From the message Y that Bob receives, how much information will he need to determine the message X that Alice sent?

To analyze this, let's see the case when Alice sends $X = x$ and Bob receives $Y = y$. The probability that this happens is the joint probability p_{xy}. Then, the probability that Bob receives the message $Y = y$, considering all the possible messages sent by Alice is

$$p_y = \sum_x p_{xy}. \qquad (10)$$

If Bob receives the message $Y = y$, the probability that Alice sent the message $X = x$ is the conditional probability denoted by $p_{x|y}$, and defined from joint probability, known as Bayes rule [12].

$$p_{xy} = p_y p_{x|y} = p_x p_{y|x}. \qquad (11)$$

On the other hand, from Bob's point of view, once he receives $Y = y$, the uncertainty about the Alice's message X, will be the entropy of X given the message $Y = y$. This is the Shannon entropy of the conditional probability $p_{x|y}$

$$H(X|Y = y) = -\sum_x p_{x|y} \log p_{x|y}. \qquad (12)$$

Using this, finally we can define the conditional entropy.

Definition 2.4 *The classical conditional entropy of random variable X given that the value of random variable Y is known, is defined as the weighted average of every possible outcome of message $Y = y$, this is*

$$H(X|Y) = \sum_y p_y H(X|Y = y). \qquad (13)$$

Replacing equation (12) and using Bayes rule (equation (11)) in definition of conditional entropy (equation (13)) we get

$$H(X|Y) = -\sum_{xy} p_{xy} \log p_{xy} + \sum_{xy} p_{xy} \log p_y, \qquad (14)$$

and given that $p_y = \sum_x p_{xy}$. Then classical conditional entropy can be written as

$$H(X|Y) = H(X, Y) - H(Y). \qquad (15)$$

This result gives us a better insight of classical conditional entropy, since the right side of the equality tells us that conditional entropy quantifies the remaining uncertainty about the composite random variable (X, Y) after subtracting the uncertainty about Y.

2.2. Quantum entropy

In the quantum case, an analogous experiment of tossing a coin is the measurement of the spin of an isolated
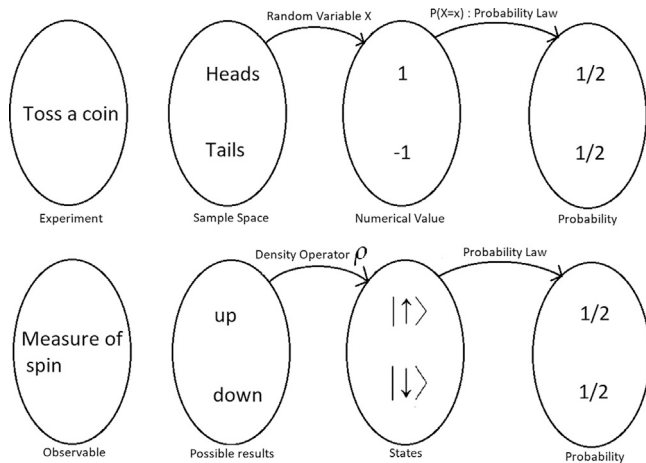
Figure 1: Analogous experiment in the classical case and in the quantum case show us the analogy between classical entropy and quantum entropy. The spin, as it have two eigenvectors, it have two possible results, for this reason is considered as a quantum coin. **(a)** Experiment of tossing a coin. The classical entropy $H(X)$ quantifies our uncertainty about whether the result will be 1 (heads) or -1 (tails). **(b)** Experiment of observing the spin of a particle. The quantum entropy $S(\rho)$ quantifies our uncertainty about whether the result will be $|\uparrow\rangle$ (up) or $|\downarrow\rangle$ (down).

electron, which is a fermion[2], described by the state $|\psi\rangle = \frac{|\uparrow\rangle_Z}{\sqrt{2}} + \frac{|\downarrow\rangle_Z}{\sqrt{2}}$. Due to the quantum nature of the spin, we will not know a priori the outcome of the experiment. Quantum mechanics only provides us a probability distribution, which is modeled by the density operator ρ. In our example, state $|\psi\rangle$, we only have the following two possible results: the outcome "spin up" with probability 1/2 and the outcome "spin down" with probability 1/2.

Here, quantum entropy will quantify the uncertainty about the outcome of this quantum observation. In other words, quantum entropy quantifies our lack of information to determine what result we will get after performing a measurement of a certain observable. A scheme of this process is draw in Figure 1b.

To formalize these ideas within the framework of quantum mechanics and in order to have a well-defined von Neumann entropy, the formalism of the density operator is necessary. The density operator ρ acts on a Hilbert space H and belongs to the space $D(\mathcal{H})$, where $D(\mathcal{H})$ is the set of operators that satisfies the following two conditions (for further reading see ref. [13, 14]).

- The density operator $\rho \in D(\mathcal{H})$ is a positive semi-definite operator, i.e.:

$$\langle\phi|\rho|\phi\rangle \geq 0, \qquad (16)$$

where $|\phi\rangle$ is an arbitrary vector in state space.

---

[2] A fermion is a particle with half-integer spin and that satisfies the Pauli exclusion principle.

- The density operator $\rho \in D(\mathcal{H})$ has trace equal to one, i.e.:

$$\mathrm{Tr}(\rho) = 1. \qquad (17)$$

As the density operator is a positive operator (it implies that density operators are hermitian see ref. [14], page 71) and its trace is unitary, we can write

$$\rho = \sum_x p_x |x\rangle\langle x|. \qquad (18)$$

Here the set $\{p_x\}$ are the eigenvalues of ρ and $\{|x\rangle\}$ its eigenvectors, which form an orthonormal basis.

The fact that density operators are positive operators implies $p_x \geq 0$ and due to its trace is unitary we have that $\sum_x p_x = 1$. Hence these density operators can be seen as a generalization of random variables and probability distributions within quantum physics. Therefore, the quantum entropy first introduced by von Neumann [15] is defined as follows.

Definition 2.5 *The quantum entropy $S(\rho)$ of a system $\mathcal{H}$ described by the density operator $\rho \in D(\mathcal{H})$ is*

$$S(\rho) = -\mathrm{Tr}(\rho \log_2 \rho) = -\sum_x p_x \log_2 p_x, \qquad (19)$$

where Tr is the trace function and the set $\{p_x\}$ are the eigenvalues of ρ.

From this, it is easy to see that the von Neumann entropy $S(\rho)$ is equal to the Shannon entropy of the probability distribution obtained from the set of eigenvalues of ρ, i.e.,

$$S(\rho) = H(\{p_x\}). \qquad (20)$$

As a reminder, the von Neumann entropy quantifies the uncertainty about what outcome we would get as a result of performing a measurement in a quantum system.

It is worth saying that we are considering the basis 2 for logarithms in both definitions of classical and quantum entropy (equations (6) and (19) respectively). This is because within the framework of Information Theory, we usually work with bits (or qubits in the quantum case), that is why it is necessary consider the base 2. Henceforth, by convention, we will not write the base 2, i.e., $\log_2 := \log$.

3. Correlations

In information theory, two random variables A and B, will be correlated if we can extract information about one random variable by observing the other. In other words, correlation implies that we can obtain information about one variable, just by making measurements on the other. For example, suppose that we have a system composed of two particles and the total charge is zero. If we separate one particle very far from the other,

it is enough to observe the charge of a single particle to know with certainty the charge of the other particle, let's say, we obtain that the charge of one particle is positive, so we know that the charge of the other particle will necessarily be negative, in order to satisfy the condition that the total charge is zero.

Using the idea of information entropy, if A and B are correlated then the reduction of uncertainty on A will allows us to reduce the uncertainty about B. Given that correlations are intimately related to the information about a random variable or a quantum system, we will study correlations using the framework of entropy of classical information introduced by Claude Shannon, and for quantum systems we will use the von Neumann entropy.

We are considering two kinds of correlations, classical and quantum correlations. A mathematical tool that allows us to quantify the correlations between two random variables is the Mutual Information [2]. In spite of mutual information being originally a measure of classical correlations, a correct generalization of this quantity will allows us to quantify total correlations [16, 17].

3.1. Classical mutual information

Classical Mutual information has a lot of equivalent definitions. Here we will only show two of them. The first one is defined as follows:

Definition 3.1 *The Classical Mutual Information I of two random variables A and B is*

$$I(A,B) = H(A) + H(B) - H(A,B), \qquad (21)$$

where $H(A,B)$ is the joint entropy.

This definition can be seen graphically in the Venn diagram (Figure 2). Notice that the mutual information $I(A,B)$ corresponds to the intersection of the information in A with the information in B. In order to understand the mutual information I, it will be necessary interpret the terms in the equality (equation 21).

The joint entropy $H(A,B)$ between two random variables can be interpreted as the lack of information about the mixture of A and B. In this definition of mutual information I (equation (21)), on the right side of the equality, we have the sum of uncertainty about random variable A plus the uncertainty about B, minus the joint entropy of A and B. This subtraction can be interpreted as the information gained if we would know the random variable (A,B). Therefore the mutual information quantifies the remain uncertainty about random variable A and B if we do not take account the uncertainty about the mixture random variable (A,B). In other words, mutual information I quantifies the shared information between random variables A and B, this shared information can be seen as classical correlations.

For the second equivalent definition of classical mutual information we will use another name (we will use a
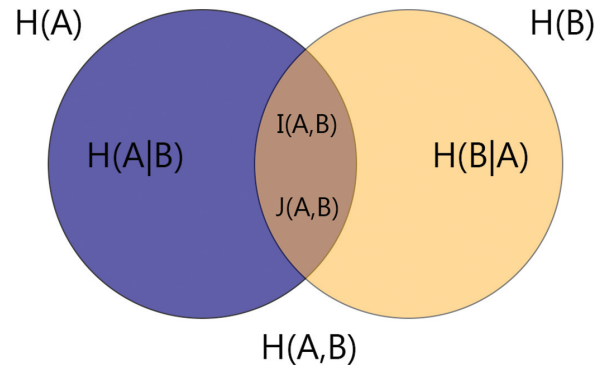

Figure 2: Venn diagram showing relationships between various information measures associated with correlated random variables A and B. The intersection between the circle of individual entropy $H(A)$ and the circle of individual entropy $H(B)$ shows the equivalence between classical mutual information $I(A,B)$ and classical conditional mutual information $J(A,B)$. The joint entropy $H(A,B)$ is the junction of entropy of random variables A and B.

different letter) in order to distinguish the first definition (equation (21)) from this second one.[3]

Definition 3.2 *The Classical Conditional Mutual Information J of two random variables A and B is*

$$J(A,B) = H(A) - H(A|B), \qquad (22)$$

where $H(A|B)$ is the conditional entropy.

Here, as we have shown in section (2.1.1), the conditional entropy $H(A|B)$ quantifies the lack of information about the random variable A when random variable B is known. And replacing the alternative form of conditional entropy (equation (15)), i.e., H(A|B) = H(A,B) − H(B) in equation (22), we obtain that conditional mutual information J is

$$J(A,B) = H(A) - H(A,B) + H(B), \qquad (23)$$

and comparing with the first definition of mutual information (equation (21)) we get the following important equivalence

$$I(A,B) = J(A,B). \qquad (24)$$

It tells us that the definition of mutual information I (def. 3.1) is equivalent to the definition of conditional mutual information J (def. 3.2). In the next part we will see that this will not happen in the quantum case. In an ilustrative way, we can see better the equivalence between I and J Eq. (24), in the Venn diagram (Figure 2).

Therefore, there are two equivalent ways of measuring classical correlations. The first one (definition 3.1)

---

[3] The reason for distinguishing one from the other will be seen later, in the quantum case.

measures the classical correlations based on the difference between the sum of the local entropies and the total entropy; the second one (definition 3.2) measures the classical correlations between two random variables based on how much information we can obtain about one random variable by extracting information[4] from the other. It is remarkable that while for I we do not have to know anything about the random variables, for J we do have to know one of them.

3.2. Quantum mutual information

Now we will generalize both definitions for the classical mutual information I and J to the quantum case. In the previous section we worked with events relates to classical discrete random variables. Here we will work with a composite quantum system $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ composed of two subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$, such that independent measurements can be made on either part. As we did for quantum entropy (section 2.2), it is necessary to generalize the classical random variables A and B to density operators acting on the Hilbert space of the system and belonging to the space of positive and unitary trace operators $D(\mathcal{H})$, this is

$$A \to \rho^A \in D(\mathcal{H}_A),$$
$$B \to \rho^B \in D(\mathcal{H}_B), \qquad (25)$$
$$(A, B) \to \rho^{AB} \in D(\mathcal{H}_{AB}).$$

In the classical case, as we have seen in equation (8), from the joint probability distribution p_{xy}, the marginal distributions can be obtained as a summation over the variables x or y. The corresponding operation for density matrices is the partial trace (see ref. [14] section 2.4.3), that is

$$\rho^A = \mathrm{Tr}_B \, \rho^{AB} = \sum_j \langle \psi_j | \, \rho^{AB} \, | \psi_j \rangle_B \,,$$
$$\rho^B = \mathrm{Tr}_A \, \rho^{AB} = \sum_i \langle \phi_i | \, \rho^{AB} \, | \phi_i \rangle_A \,, \qquad (26)$$

where $|\psi_j\rangle_B \in \mathcal{H}_B$, $|\phi_i\rangle_A \in \mathcal{H}_A$ and ρ^A (or ρ^B) is the reduced state also known as the marginal state of ρ^{AB} on subsystem A (or B). Therefore, the quantum entropy of ρ^A is

$$S(\rho^A) = S(A) = -\mathrm{Tr}(\rho^A \ln \rho^A), \qquad (27)$$

where we used the notation $S(\rho^A) := S(A)$. The quantum entropy ρ^B is defined in the same way.

Now we want to generalize the classical mutual information I defined in equation (21) to the quantum case. This generalizations seems to be natural, i.e., we simply replaced the Shannon entropy $H()$ for the quantum entropy $S()$:

Definition 3.3 *The quantum Mutual Information $\mathcal{I}$ between two quantum systems A and B, described respectively by the density operators $\rho^A \in D(\mathcal{H}_A)$ and $\rho^B \in D(\mathcal{H}_B)$ is defined as*

$$\mathcal{I}(A, B) := S(A) + S(B) - S(A, B), \qquad (28)$$

where $S(A, B) = -\mathrm{Tr}(\rho^{AB} \ln \rho^{AB})$ is the quantum entropy of the composite system described by $\rho^{AB} \in D(\mathcal{H})$.

Here the quantum mutual information $\mathcal{I}$ is quantifying the total correlations between systems A and B. These definitions tells us that if the total correlations are not zero, i.e., $\mathcal{I} \neq 0$, then the density matrix ρ^{AB} for the entire system is not equal to the tensor product $\rho^A \otimes \rho^B$ of the reduced density matrices. This means that the correlations between system A and B are not included in $\rho^A \otimes \rho^B$ [8]. We can formalize this ideas following the next theorem.

Theorem 3.4 *The quantum mutual information $\mathcal{I}$ can be written in the following form*

$$\mathcal{I}(\rho^{AB}) = S(\rho^{AB} || \rho^A \otimes \rho^B), \qquad (29)$$

where $S(\rho || \sigma) = \mathrm{Tr}(\rho \log_2 \rho) - \mathrm{Tr}(\rho \log_2 \sigma)$ is the quantum relative entropy.

The quantum relative entropy measures how close are two density operators. Therefore, from equation (29) we can interpret that the correlations between systems A and B can be measured by measuring the distance between ρ^{AB} and $\rho^A \otimes \rho^B$.

The problem arises when we want to generalize the second definition of conditional mutual information $\mathcal{J}$ (equation (22)) to the quantum level. This is because, if we do the same substitution for conditional entropy, as before, i.e. substitute the classical entropy for the quantum entropy in equation (15), the quantum conditional entropy $S(A|B)$ would be

$$S(A|B) = S(A, B) - S(B), \qquad (30)$$

and, if we consider a pure entangled state, for example, the bell state $|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, the density operator would be $\rho^{AB} = |\beta_{00}\rangle \langle \beta_{00}|$. The quantum entropy of the composite system is

$$S(A, B) = 0, \qquad (31)$$

while the quantum entropy of the subsystem B is

$$S(B) = \frac{1}{2} \ln 2. \qquad (32)$$

Replacing this results in Eq. (30), we can see that the a priori[5] generalization could lead to a negative quantum

---

[4] Henceforth we will use observation, measuring or extract information as synonyms.

[5] It is because we just replace the classical entropy for the quantum entropy, without taking into account the quantum phenomena.

conditional entropy. In our example of the particular case of a pure entangled state, we get that

$$S(A|B) = -\frac{1}{2}\ln 2 < 0. \qquad (33)$$

Due to entropy quantifies the uncertainty (or the lack of information) about a state, a negative entropy is difficult to interpret.[6] In other words, the fact that $S(A) = 0$, means that we do not have uncertainty (or we have all the information) about the state A, so it is hard to see how this quantity could be negative. Therefore, due to the necessity of quantum conditional entropy in the generalization of quantum conditional mutual information $\mathcal{J}$, we have to find another way to define the quantum conditional entropy in a consistent form.

3.2.1. Quantum conditional entropy

Olivier and Zurek [16] proposed the following way to generalize the quantum conditional entropy. Since the classical conditional entropy $H(A|B)$ (defined in equation (13)) quantifies the remain uncertainty about the variable A when B is known. Within the quantum theory framework, stating that B is known is ambiguous, since to know the state of the system we have to perform a measurement, which implies that we have to specify a set of measurement operators on the state space of the system. Therefore, a consistent definition of quantum conditional entropy requires us to specify a set of measurements performed in one subsystem. For example, we will choose the set of measurements $\{P_j^B\}$ performed on subsystem B. Here the index j serves to specify the j-th outcome of the measurement as a result of the measurement P_j. Following the measurement postulate of quantum mechanics [14, 19, 20], the state of the composite system ρ^{AB}, once the measurement operator $\mathbb{I}_A \otimes P_j^B$ is performed, will collapse to the state

$$\rho^{A|P_j^B} = \frac{(\mathbb{I}_A \otimes P_j^B)\rho^{AB}(\mathbb{I}_A \otimes P_j^B)}{\mathrm{Tr}_{AB}[(\mathbb{I}_A \otimes P_j^B)\rho^{AB}]}. \qquad (34)$$

Here $\mathbb{I}_A$ is the identity operator in the state space of system A and the notation $\rho^{A|P_j^B}$ represents the state of system A due the measurement operator $\mathbb{I}_A \otimes P_j^B$.

The probability of observing the outcome j-th is:

$$p_j = \mathrm{Tr}_{AB}(P_j^{AB}\rho^{AB}). \qquad (35)$$

Therefore, the von Neumann entropy of this state $S(\rho^{A|P_j^B})$, can be interpreted as the uncertainty about the system A once the measurement operator P_j^B is performed. However, we are interested in a complete set of measurement operators $\{P_j^B\}$. Thus, to construct the conditional entropy we have to weight the entropies $S(\rho^{A|P_j^B})$ with their respective probabilities p_j.

---

[6] N. Cerf [18] interpret the negative entropy as virtual information. He and other people use the idea of negative entropy to do calculations in the information on black holes.

Definition 3.5 *The quantum conditional entropy of system A given the set of measurements $\{P_j^B\}$ performed on system B, is defined as*

$$S(A|\{P_j^B\}) = \sum_j p_j S(\rho^{A|P_j^B}), \qquad (36)$$

where $\rho^{A|P_j^B}$ is the state[7] to which the composite system collapses after the measurement P_j^B.

There are infinitely many sets of measurements we can perform on B. For the sake of simplicity we will consider the set of one-dimensional projectors.[8] As we want to learn as much as possible about system A by measuring system B, then we will choose the set that makes $S(A|\{P_j^B\})$ minimum. It is clear that the upper bound on this quantum conditional entropy is $S(A)$, while the lower bound is zero.

Finally, we can give the quantum generalization of the conditional mutual information $\mathcal{J}$ as follows.

Definition 3.6 *The quantum Conditional Mutual Information $\mathcal{J}$ between two quantum systems A and B, described respectively by the density operators $\rho^A \in D(\mathcal{H}_A)$ and $\rho^B \in D(\mathcal{H}_B)$ is defined as*

$$\mathcal{J}(A,B)_{\{P_j^B\}} := S(A) - S(A|\{P_j^B\}), \qquad (37)$$

where $\{P_j^B\}$ is a complete set of measurements perform on system B.

This quantity represents the information gained about subsystem A as a result of the set of measurements $\{P_j^B\}$ performed on subsystem B. Recall that the measurements that we are considering are one-dimensional projector operators. Therefore, obtaining information from A only performing measurements on B tells us that there exists correlations. Otherwise, if there are no correlations, then the conditional entropy would be $S(A|\{P_j^B\}) = S(A)$ and therefore $\mathcal{J} = 0$. That is why quantum conditional mutual information serves to quantify correlations, but what kind of correlations? Since definition of $\mathcal{J}$ (equation (37)) needs us to perform a complete measurement on B, due to the nature of the measurement process, the state ρ^{AB} of system AB will collapse to a quantum classical state where the marginal density operator on B will be a classical state in the sense that it will not be disturb by certain local measurements in system B (see ref. [21, 22]). It implies that the subsystem B will have classical nature, hence the correlations between A and B will be classical. Therefore the quantum conditional mutual information $\mathcal{J}$ quantifies the classical correlations between quantum systems A and B.

---

[7] The state ρ^{AB} of the composite system will collapse to the state $\rho^{A|P_j^B} = \frac{P_j^B \rho^{AB} P_j^B}{p_j}$, with probability $p_j = \mathrm{Tr}_{AB}(P_j^B \rho^{AB})$ when the measurement P_j^B is performed.

[8] The one-dimensional projectors are those operators with only one nonzero eigenvalue.

4. Final Considerations

In order to have a better understanding of the difference between quantum mutual information and quantum conditional mutual information we will show an example. Suppose that we have the following quantum state ρ^{AB} describing a system composed of two subsystems A and B and that depends on the variable $c \in [0, 1]$:

$$\rho^{AB} = \frac{(1+c)}{4}\mathbb{I} - c\,|\beta_{11}\rangle\,\langle\beta_{11}|, \qquad (38)$$

where $\mathbb{I}$ is the identity operator in the state space $\mathcal{H}_A \otimes \mathcal{H}_B$, i.e., $\mathbb{I} := \mathbb{I}_A \otimes \mathbb{I}_B$ and $|\beta_{11}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$ is a bell state (see ref. [14], sec. 1.3.6). If we perform the calculation of mutual information $\mathcal{I}$ from equation (3.3) and conditional mutual information $\mathcal{J}$ from equation (3.6) using the state ρ^{AB} we will get respectively (see ref. [23]):

$$\mathcal{I}(\rho^{AB}) = \frac{(1-3c)}{4}\log(1-3c) + \frac{3(1+c)}{4}\log(1+c). \qquad (39)$$

$$\mathcal{J}(\rho^{AB}) = \frac{(1-c)}{2}\log(1-c) + \frac{(1+c)}{2}\log(1+c). \qquad (40)$$

From here is easy to see that, apart from the trivial case $c = 0$, these quantities are different, i.e., $\mathcal{I} \neq \mathcal{J}$. Therefore although at the classical case both definitions of mutual information I and conditional mutual information J are equal (Equation (24)), i.e.:

$$I(A, B) = J(A, B), \qquad (41)$$

in the quantum case, in general, they are no longer equivalent, i.e

$$\mathcal{I}(A, B) \neq \mathcal{J}(A, B). \qquad (42)$$

This will become clear seeing the following upper bounds. For the quantum mutual information $\mathcal{I}$ we have the Araki-Lieb inequality [6]:

$$\mathcal{I}(A, B) \leq 2\min\{S(A), S(B)\}, \qquad (43)$$

while for the quantum mutual information $\mathcal{J}$ is easy to see that

$$\mathcal{J}(A, B) \leq S(A). \qquad (44)$$

Therefore if $\mathcal{I}$ and $\mathcal{J}$ reach their upper bounds, i.e., $\mathcal{I}(A, B) = 2S(A)$ and $\mathcal{J}(A, B) = S(A)$. Then, the difference will be

$$\mathcal{I}(A, B) - \mathcal{J}(A, B) \neq 0, \qquad (45)$$

which is the same than Eq. (42).

The natural question is, what does the difference between the two quantum quantities $\mathcal{I}(A, B) - \mathcal{J}(A, B)$ signify, if anything? [24]. We have said that the Mutual Information $\mathcal{I}(A, B)$ quantifies the total correlations, and that the Mutual Information $\mathcal{J}(A, B)$ quantifies classical correlations. Therefore, under certain conditions, what remains from the subtraction of the total correlations and the classical correlations, i.e. $\mathcal{I}(A, B) - \mathcal{J}(A, B)$, will be the quantum correlations [16].

On the other hand, we know that quantum entanglement is a quantum correlation. Hence the question is that if entanglement is the only type of quantum correlation. Henderson and Vedral [17] claim that quantum mutual information $\mathcal{I}$ quantity can be expressed as a sum of classical correlations and entanglement. To prove it, they defined the classical correlations of a composite system by $C(A, B) = S(A) - S(A|B)$ and quantified entanglement using the entanglement of formation $E(A, B)$ for mixed states [25, 26]. They discovered that the sum $C + E$ is mostly smaller than $\mathcal{I}$ [17]. In other words, there is more to quantum correlations than just entanglement when it comes to mixed states. For pure states, entanglement and classical correlations are equal to each other and the sum is then exactly equal to the quantum mutual information $\mathcal{I}$, what explains why $\mathcal{I}$ is twice as big as the quantum mutual information $\mathcal{J}$ (inequalities (43) and (44)).

In conclusion, both quantum mutual information $\mathcal{I}$ and $\mathcal{J}$ allow us quantify directly the classical correlations and total correlations and indirectly (subtracting $\mathcal{I} - \mathcal{J}$) the quantum correlations. It is remarkable that entanglement is not the only type of quantum correlations.

Acknowledgements

References

[1] E. Andrade, Revista Colombiana de Filosofía de la Ciencia **17**, 34 (2017).

[2] C. Thomas and J. Thomas, *Elements of Information Theory* (Wiley-Interscience, New Jersey, 1991).

[3] C.E. Shannon, The Bell System Technical Journal **27**, 379 (1948).

[4] C.E. Shannon, The Bell System Technical Journal **27**, 623 (1948).

[5] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Champaign, 1949).

[6] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[7] C.E. Shannon, The Bell System Technical Journal **30**, 50 (1951).

[8] G. Benenti , G. Casati and G. Strini, *Principles of Quantum Computation and Information* (World Scientific, Singapore, 2007).

[9] R. Dorner and V. Vedral, International Journal of Modern Physics B **27**, 1345017 (2013).

[10] J. Maziero, Rev. Bras. Ens. Fis. **37**, 1314 (2015).

[11] M.H. DeGroot and M.J. Schervish, *Probability and Statistics* (Addison Wesley Longman, Boston, 1975).

[12] T. Bayes, Philosophical Transactions of the Royal Society of London, **53**, 370 (1763).

[13] R. Balian, *From Microphysics to Macrophysics* (Springer, Paris, 1991).

[14] M. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[15] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).

[16] H. Olliver and W. Zurek, Phys. Rev. Lett. **88**, 017901 (2001).

[17] L. Henderson and V. Vedral, Journal of Physics A: Mathematical and General **34**, 6899 (2001).

[18] N. Cerf and C. Adami, Physical Review Letters **79**, 5194 (1997).

[19] A. Peres, *Quantum Theory: Concepts and methods* (Kluwer Academic Publishers, New York, 1995).

[20] V. Vedral, *Introduction to Quantum Information Science* (Oxford University Press, New York, 2006).

[21] A. Streltsov, *Quantum Correlations Beyond Entanglement* (Springer, Barcelona, 2015).

[22] M. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen(De), Ujjwal Sen and B. Synak-Radtke, Phys. Rev. A **71**, 062307 (2005).

[23] S. Luo, Phys. Rev. A **7**, 042303 (2008).

[24] G. Lindblad, Communications in Mathematical Physics **33**, 305 (1973).

[25] V. Vedral, Phys. Rev. Lett. **78**, 2275 (1997).

[26] W. Wootters, Phys. Rev. Lett. **80**, 2245 (1998).