

# PI-Measure: A Set-Theoretic View of Partial Information

Litao Zhou, Hongbo Yang, Shiwen Dong

August 13, 2020

## 1 Introduction

Developing from investigation in physics, engineering and mathematics, information theory originated from Shannon's pioneering research of the reliability and coding in communication system. It has successfully become a universal tool for analyzing complex systems, and has been applied in many fields such as genetics, physics, neuroscience, machine learning and so on. Shannon's mutual information is by far the most widely used concept from the information theory [1], and Shannon's information measures refer to entropy, conditional entropy, mutual information, and conditional mutual information, which are the most significant measures of information in information theory.

The relations between the entropies, conditional entropies and mutual information for two random variables  $X$  and  $Y$  can be summarized by a variation of the Venn diagram. With the help of such diagram, the relations can be depicted intuitively and vividly, thus reducing the difficulty of readers' understanding of the knowledge. This one-to-one correspondence between Shannon's information measures and set theory is not just a coincidence for two random variables. Actually, Shannon's information measures for any  $n \geq 2$  random variables all have a set-theoretic structure.

Using a few basic concepts in measure theory such as fields, atoms and signed measure  $\mu$ , a theory "I-measure" which establishes a one-to-one correspondence between Shannon's information measures and set theory in full generality has been developed [2, 3]. Then manipulations of Shannon's information measures can be viewed as set operations, thus allowing the rich suite of tools in set theory to be used in information theory. Moreover, the structure of Shannon's information measures can easily be visualized by means of an information diagram if four or fewer random variables are involved. The use of information diagrams simplifies many difficult proofs in information theory problems. More importantly, these results, which may be difficult to discover in the first place, can easily be obtained by inspection of an information diagram.

Until now, most of the work of information theory only deals with the simplest case: the information about another variable provided by one variable. By contrast, most challenging scientific problems, such as many-body problems in physics [4], and population coding in neuroscience [5, 6], involve understanding the structure of interactions between three or more variables. To generalize information theory to multivariate interactions, interaction information was proposed as a measure of the amount of information bound in a set of variables, rather than the amount of information existing in any subset of those variables. So entropy and mutual information correspond to the first- and second-order interaction information, while interaction information and its third-order, fourth-order and higher-order variants provide a method to represent the structure of multivariate information. However, the widespread use of interactive information is largely hindered by "odd" and "unfortunate" attributes. For three or more variables, interaction information can be negative, whose meaning is usually unclear for information as it is commonly understood.

Then a new perspective on the structure of multivariate information has been formulated.[7] Considering the general structure of the information that a set of sources provide about a given variable, a new definition of redundancy was proposed, which can be used to exhaustively decompose the Shannon information in a multivariate system into partial information atoms.

Here we use the above-mentioned method of I-measure, aiming to establish a one-to-one correspondence between partial information measures and the set theory. With this correspondence, manipulations of partial information measures can be viewed as set operations. What's more, the structure of partial information measures can be easily visualized by a partial information diagram. And we hope that this can simplify the deduction and proofs in information theory problems.

## 2 I-Measure

We have already known that the Venn diagrams can give us a clear picture of the relations among sets. In order to construct similar information diagrams to facilitate our analysis of information, we need to introduce a method to connect Shannon's information measure with things in set theory. We introduce a few basic concepts in measure theory which will be used subsequently. These concepts will be illustrated by simple examples.

**Definition 1** (Field). *The field  $\mathcal{F}_n$  generated by sets  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$  is the collection of sets which can be obtained by any sequence of usual set operations (union, intersection, complement, and difference) on  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$*

**Definition 2** (Atom). *The atoms of  $\mathcal{F}_n$  are sets of the form  $\cap_{i=1}^n Y_i$ , where  $Y_i$  is either  $\tilde{X}_i$  or  $\tilde{X}_i^c$ , the complement of  $\tilde{X}_i$*

**Definition 3** (Signed Measure). *A real function  $\mu$  defined on  $\mathcal{F}_n$  is called a signed measure if it is set-additive, i.e., for disjoint  $A$  and  $B$  in  $\mathcal{F}_n$*

$$\mu(A \cup B) = \mu(A) + \mu(B) \quad (1)$$

For a signed measure  $\mu$ , we have

$$\mu(\emptyset) = 0 \quad (2)$$

which can be seen as follows. For any  $A$  in  $\mathcal{F}_n$

$$\mu(A) = \mu(A \cup \emptyset) = \mu(A) + \mu(\emptyset) \quad (3)$$

by set-additivity because  $A$  and  $\emptyset$  are disjoint. A signed measure  $\mu$  on  $\mathcal{F}_n$  is completely specified by its values on the atoms of  $\mathcal{F}_n$ . The values of  $\mu$  on the other sets in  $\mathcal{F}_n$  can be obtained via set-additivity.

Until now, we have found a practical method to measure a set. We'll see how we can introduce a measure for the information in section 2.1.

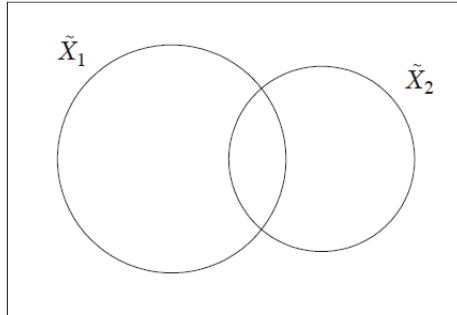


Figure 1: The Venn diagram for  $\tilde{X}_1$  and  $\tilde{X}_2$

### 2.1 I-Measure between 2 R.V.s

To start with, we first formulate the one-to-one correspondence between Shannon's information measures and set theory for two random variables. For random variables  $X_1$  and  $X_2$ , let  $\tilde{X}_1$  and  $\tilde{X}_2$  be sets respectively corresponding to  $X_1$  and  $X_2$ . The sets  $\tilde{X}_1$  and  $\tilde{X}_2$  generates the field  $\mathcal{F}_n$  whose atoms are

$$\tilde{X}_1 \cap \tilde{X}_2, \tilde{X}_1^c \cap \tilde{X}_2, \tilde{X}_1 \cap \tilde{X}_2^c, \tilde{X}_1^c \cap \tilde{X}_2^c$$

Their relationship is shown by a Venn diagram in Figure 1.

In our formulation, we set the universal set  $\Omega$  to  $\tilde{X}_1 \cup \tilde{X}_2$ . With this choice of  $\Omega$ , the Venn diagram for  $\tilde{X}_1$  and  $\tilde{X}_2$  is represented by the diagram in Figure 2. For simplicity, the sets  $\tilde{X}_1$  and  $\tilde{X}_2$  are respectively labeled by  $X_1$  and  $X_2$  in the diagram. We call this the information diagram for the random variables  $X_1$  and  $X_2$ . In

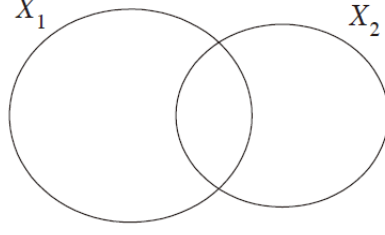


Figure 2: The information diagram for  $X_1$  and  $X_2$

this diagram, the universal set, which is the union of  $\tilde{X}_1$  and  $\tilde{X}_2$  is not shown explicitly just as in a usual Venn diagram. Note that with our choice of the universal set, the atom  $\tilde{X}_1^c \cap \tilde{X}_2^c$  degenerates to the empty set.

From what we know about random variables  $X_1$  and  $X_2$ , we can explicit the Shannon's information measures

$$H(X_1), H(X_2), H(X_1|X_2), H(X_2|X_1), H(X_1, X_2), I(X_1; X_2)$$

Writing  $A \cap B^c$  as  $A - B$ , we now define a signed measure  $\mu^*$  by

$$\mu^*(\tilde{X}_1 - \tilde{X}_2) = H(X_1|X_2) \quad (4)$$

$$\mu^*(\tilde{X}_2 - \tilde{X}_1) = H(X_2|X_1) \quad (5)$$

and

$$\mu^*(\tilde{X}_1 \cap \tilde{X}_2) = I(X_1; X_2) \quad (6)$$

These are the values of  $\mu^*$  on the nonempty atoms of  $\mathcal{F}_2$  (i.e., atoms of  $\mathcal{F}_2$  other than  $\tilde{X}_1^c \cap \tilde{X}_2^c$ ). The values of  $\mu^*$  on the other sets on  $\mathcal{F}_2$  can be obtained via set-additivity. In particular, we can easily formulate these following relations

$$\mu^*(\tilde{X}_1 \cup \tilde{X}_2) = H(X_1, X_2) \quad (7)$$

$$\mu^*(\tilde{X}_1) = H(X_1) \quad (8)$$

and

$$\mu^*(\tilde{X}_2) = H(X_2) \quad (9)$$

For example,  $\mu^*(\tilde{X}_1 \cup \tilde{X}_2)$  can be calculated by

$$\mu^*(\tilde{X}_1 \cup \tilde{X}_2) \quad (10)$$

$$= \mu^*(\tilde{X}_1 - \tilde{X}_2) + \mu^*(\tilde{X}_2 - \tilde{X}_1) + \mu^*(\tilde{X}_1 \cap \tilde{X}_2) \quad (11)$$

$$= H(X_1|X_2) + H(X_2|X_1) + I(X_1; X_2) \quad (12)$$

$$= H(X_1, X_2) \quad (13)$$

By now, the relationship between Shannon's information measures and signed measure (i.e. the set theory) is clear for 2 R.V.s. By the approach of sign mapping, we can change between Shannon's information measures into signed measure, and vice versa.

## 2.2 Construction of $\mu^*$

In this section, we are finding a more general conclusion, which means to find a similar relation in a higher random variable space. For  $n \geq 2$ , consider  $n$  random variables  $X_1; X_2; \dots; X_n$ . For any random variable  $X$ , let  $\tilde{X}$  be a set corresponding to  $X$ . Let

$$\mathcal{N}_n = \{1, 2, \dots, n\} \quad (14)$$

Define the universal set  $\Omega$  to be the union of the sets  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ , i.e.

$$\Omega = \bigcup_{i \in \mathcal{N}_n} \tilde{X}_i \quad (15)$$

And we use  $\mathcal{F}_n$  to denote the field generated by  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ .

The set

$$A_0 = \bigcap_{i \in \mathcal{N}_n} \tilde{X}_i^c$$

is called the empty atom of  $\mathcal{F}_n$  because

$$\bigcap_{i \in \mathcal{N}_n} \tilde{X}_i^c = \left( \bigcup_{i \in \mathcal{N}_n} \tilde{X}_i \right)^c = \Omega^c = \emptyset \quad (16)$$

All the atoms of  $\mathcal{F}_n$  other than  $A_0$  are called nonempty atoms.

To simplify notation, we will use  $X_G$  to denote  $(X_i, i \in G)$  and  $\tilde{X}_G$  to denote  $\cup_{i \in G} \tilde{X}_i$  for any nonempty subset  $G$  of  $\mathcal{N}_n$

**Theorem 4.** *Let*

$$\mathcal{B} = \left\{ \tilde{X}_G : G \text{ is a nonempty subset of } \mathcal{N}_n \right\} \quad (17)$$

*Then a signed measure  $\mu$  on  $\mathcal{F}_n$  is completely specified by  $\{\mu(B), B \in \mathcal{B}\}$  which can be any set of real numbers.*

When introducing our work of "PI-Measure", we will use a similar way as this proof to show a connection in  $\mu^*$  and decomposed mutual information, and that proof will be given.

Until now the work to expand signed measures into higher space is finished. Next we'll look at how to substitute signed measures into Shannon's information measure when  $n \geq 2$ .

We now construct the I-Measure  $\mu^*$  on  $\mathcal{F}_n$  using Theorem 4 by defining

$$\mu^* \left( \tilde{X}_G \right) = H(X_G) \quad (18)$$

for all nonempty subsets  $G$  of  $\mathcal{N}_n$ . In order for  $\mu^*$  to be meaningful, it has to be consistent with all Shannon's information measures. In that case, the following must hold for all (not necessarily disjoint) subsets  $G, G', G''$  of  $\mathcal{N}_n$  where  $G$  and  $G'$  are nonempty:

$$\mu^* \left( \tilde{X}_G \cap \tilde{X}_{G'} - \tilde{X}_{G''} \right) = I(X_G; X_{G'} | X_{G''}) \quad (19)$$

When  $G'' = \emptyset$ , (19) becomes

$$\mu^* \left( \tilde{X}_G \cap \tilde{X}_{G'} \right) = I(X_G; X_{G'}) \quad (20)$$

When  $G = G'$ , (19) becomes

$$\mu^* \left( \tilde{X}_G - \tilde{X}_{G''} \right) = H(X_G | X_{G''}) \quad (21)$$

When  $G = G'$  and  $G'' = \emptyset$ , (19) becomes

$$\mu^* \left( \tilde{X}_G \right) = H(X_G) \quad (22)$$

From these cases we know that (19) can cover all the 4 conditions that Shannon's information measure may have. Thus, we can say that Equation (19) holds, if and only if  $\mu^*$  is consistent with all Shannon's information measures.

**Theorem 5.**  *$\mu^*$  is the unique signed measure on  $\mathcal{F}_n$  which is consistent with all Shannon's information measures.*

Proof. Consider

$$\begin{aligned}
& \mu^* \left( \tilde{X}_G \cap \tilde{X}_{G'} - \tilde{X}_{G''} \right) \\
&= \mu^* \left( \tilde{X}_{G \cup G''} \right) + \mu^* \left( \tilde{X}_{G' \cup G''} \right) - \mu^* \left( \tilde{X}_{G \cup G' \cup G''} \right) - \mu^* \left( \tilde{X}_{G''} \right) \\
&= H(X_{G \cup G''}) + H(X_{G' \cup G''}) - H(X_{G \cup G' \cup G''}) - H(X_{G''}) \\
&= I(X_G; X_{G'} | X_{G''})
\end{aligned}$$

i.e.,  $\mu^*$  is consistent with all Shannon's information measures. In order that  $\mu^*$  is consistent with all Shannon's information measures, for all nonempty subsets  $G$  of  $\mathcal{N}_n$ ,  $\mu^*$  has to satisfy Equation (19), which in fact is the definition of  $\mu^*$ . Therefore,  $\mu^*$  is the unique signed measure on  $\mathcal{F}_n$  which is consistent with all Shannon's information measures.

### 2.3 Markov Chain, a Special Case

All the contents above together, give us the possibility to draw a clear picture of the relationship among Shannon's information measure of discrete random variables. (The  $n = 3$  case is shown below)

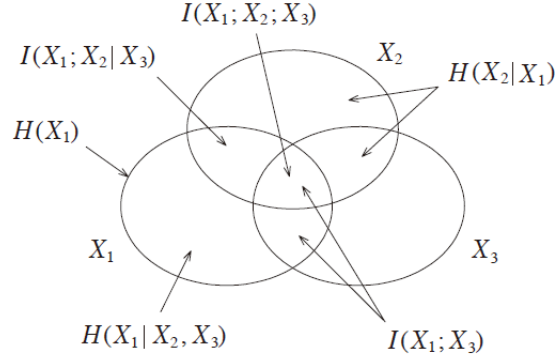


Figure 3: The information diagram for  $X_1$ ,  $X_2$  and  $X_3$

In this section, we'll focus on a special case of this relation, i.e. the Markov Chain (MC). We explore the structure of Shannon's information measures when  $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$  forms a Markov chain. To start with, we consider  $n = 3$ , i.e.,  $X_1 \rightarrow X_2 \rightarrow X_3$  forms a Markov chain. since

$$\mu^* \left( \tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \right) = I(X_1; X_3 | X_2) = 0$$

the atom  $\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3$  does not have to be displayed in an information diagram. Therefore, in constructing the information diagram, the regions representing the random variables  $X_1$ ,  $X_2$ , and  $X_3$  should overlap with each other such that the region corresponding to the atom  $\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3$  is empty, while the regions corresponding to all other nonempty atoms are nonempty. Figure 3.7 shows such a construction, in which each random variable is represented by a "mountain." From Figure 3.7, we see that  $\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3$ , as the only atom on which  $\mu^*$  may take a negative value, now becomes identical to the atom  $\tilde{X}_1 \cap \tilde{X}_3$ . Therefore, we have

$$\begin{aligned}
I(X_1; X_2; X_3) &= \mu^* \left( \tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3 \right) \\
&= \mu^* \left( \tilde{X}_1 \cap \tilde{X}_3 \right) \\
&= I(X_1; X_3) \\
&\geq 0
\end{aligned}$$

Information Diagrams for Zero Measure In Markov Chain  $X_1 \rightarrow X_2 \rightarrow X_3$ , since

$$\mu^* \left( \tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \right) = I(X_1; X_3 | X_2) = 0 \quad (23)$$

We don't have to plot that region out in the information diagram, making the diagram simpler and more useful when reasoning about Shannon information measures. And for MC with 4 or more R.V.s, we can also find out that some parts needn't to be plotted, because of the properties of Markov Chain.

Above all, we've introduced several general ideas in constructing the correspondence between set theory and Shannon information. These ideas will be revisited later, helping us find a correspondence between another type of information measure and set theory.

### 3 Decomposition of Multivariate Information

#### 3.1 The Structure of Multivariate Information

Suppose there is a random variable  $S$  whose information is provided by a random  $\mathbf{R} = (R_1, R_2, \dots, R_{n-1})$ , then we try to decompose the information into some information contributed by the subsets of  $\mathbf{R}$  individually or jointly.

Consider the case where  $\mathbf{R} = (R_1, R_2)$ . The total information  $R_1$  and  $R_2$  provided is the mutual information  $I(S; R_1, R_2)$ , and their contributions can be divided into three parts. First, the unique information, which means the information only provided by  $R_1$ , or vice versa. Second, the redundancy, which means the same or overlapping information that  $R_1$  and  $R_2$  provide. Third, the synergy, which means the information that is provided by the combination of  $R_1$  and  $R_2$  while not available from either alone.

Therefore, as is shown in Figure 4, the total information can be decomposed into three parts: the synergistic information contributed jointly by  $R_1$  and  $R_2$ , the unique information from  $R_1$  or  $R_2$ , and the redundant information shared by  $R_1$  and  $R_2$ . So we can identify them as the basic atoms of multivariate information.

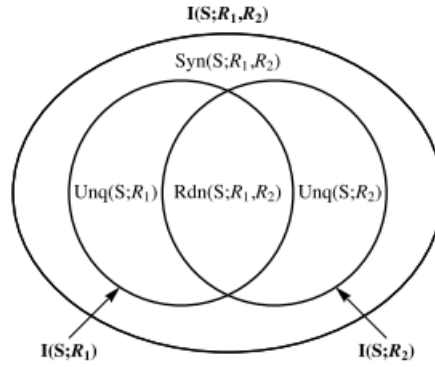


Figure 4: Structure of multivariate information for 3 variables

Actually, the unique information can be considered as a deformation of synergy or redundancy, so we can say that the multivariate information constitutes of redundancy and synergy.

#### 3.2 Redundancy Measure

To formally measure the redundancy between multiple collections of sources, we now introduce a redundancy measure  $I_{\min}$ . Let  $A_1, A_2, \dots, A_k$  be nonempty and potentially overlapping subsets of  $\mathbf{R}$ , which we call sources. We introduce  $I_{\min}$  to measure the information that all sources provide about  $S$ .

While the specific information  $I(S = s; \mathbf{A})$  quantifies the information associated with a particular outcome of  $S$ ,  $I_{\min}$  describes all sources offer to  $S$ , or we say "redundancy". Various definitions of specific information have been proposed to quantify different relationships between  $S$  and  $\mathbf{A}$  but for our purposes the most useful is

$$I(S = s; \mathbf{A}) = \sum_{\mathbf{a}} p(\mathbf{a}|s) \left[ \log \frac{1}{p(s)} - \log \frac{1}{p(s|\mathbf{a})} \right] \quad (24)$$

To better introduce the meaning of this definition, we use  $\frac{1}{p(s)}$  instead of  $p(s)$ . And the former one is called the surprise of  $s$ , so  $I(S = s; \mathbf{A})$  is the average reduction in surprise of  $s$  given knowledge of  $\mathbf{A}$ . In other words,  $I(S = s; \mathbf{A})$  quantifies the information that  $\mathbf{A}$  provides about each particular outcome  $s \in S$ , while  $I(S; \mathbf{A})$  is

the expected value of this quantity over all outcomes of  $S$ . Given these considerations, a natural measure of redundancy is the expected value of the minimum information that any source provides about each outcome of  $S$ .

**Definition 6** (Redundancy Measure).

$$I_{\min}(S; \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k\}) = \sum_s p(s) \min_{\mathbf{A}_i} I(S = s; \mathbf{A}_i) \quad (25)$$

where the domain of  $I_{\min}$   $\mathcal{A}(\mathbf{R}) = \{\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbf{R})) : \forall \mathbf{A}_i, \mathbf{A}_j \in \alpha, \mathbf{A}_i \not\subseteq \mathbf{A}_j\}$

$I_{\min}$  captures the idea that redundancy is the information common to all sources (the minimum information that any source provides), while taking into account that sources may provide information about different outcomes of  $S$ . Note that, like the mutual information,  $I_{\min}$  is also an expected value of specific information terms. And similarly to  $I$ ,  $I_{\min}$  also has several useful properties that can be easily proved from its definition.

- $I_{\min}$  is non-negative
- $I_{\min}$  is less than or equal to  $I(S; \mathbf{A}_i)$  for all  $\mathbf{A}_i$  's
- For a given source  $\mathbf{A}$ , the amount of information redundant with  $\mathbf{A}$  is maximal for  $I_{\min}(S; \{\mathbf{A}\}) = I(S; \mathbf{A})$ .

These properties can confirm  $I_{\min}$  with its interpretation as a measure of redundancy.

### 3.3 Partial Information Decomposition

Note that  $I_{\min}$  can measure redundancy between collections of sources like  $\{\{R_1\}, \{R_2, R_3\}\}$  denoted as  $\{1\}\{23\}$ . A natural observation is that, for any source sets in one collection, if its subsets can always be found at another collection, then it must measure a greater redundancy over the other collection. Therefore, we can define a partial order over the elements of  $\mathcal{A}(\mathbf{R})$  such that one is considered to precede another if and only if the latter provides any redundant information that the former provides. Formally,

$$\forall \alpha, \beta \in \mathcal{A}(\mathbf{R}), \alpha \preceq \beta \Leftrightarrow \forall \mathbf{B} \in \beta, \exists \mathbf{A} \in \alpha, \mathbf{A} \subseteq \mathbf{B} \quad (26)$$

And from this partial order, a lattice can be illustrated as follow, which gives us a clear picture of the partial order among collections of sources.(shown in Figure 5)

The redundant information associated with each node of the redundancy lattice includes, but is not limited to, the redundant information provided by all nodes lower in the lattice. Thus moving from node to node up the lattice,  $I_{\min}$  can be thought of as a kind of "cumulative information function," effectively integrating the information provided by increasingly inclusive collections of sources.

The observation above implies that the nodes in the lattice actually would decompose redundancy measure into several non-intersecting fragments and derive an inverse of  $I_{\min}$  called the *partial information function* (PI-function). Whereas  $I_{\min}$  quantifies cumulative information, the PI-function measures the partial information contributed uniquely by each particular collection of sources. This partial information will form the atoms into which we decompose the total information that  $\mathbf{R}$  provides about  $S$ . Formally, we define these components with the partial order defined in Equation 26.

**Definition 7** (Partial Information). For a collection of sources  $\alpha \in \mathcal{A}(\mathbf{R})$ , the PI-function, denoted  $\Pi_{\mathbf{R}}$ , is defined implicitly by

$$I_{\min}(S; \alpha) = \sum_{\beta \preceq \alpha} \Pi_{\mathbf{R}}(S; \beta) \quad (27)$$

The relationship between these measures can be shown in a partial information (PI) diagram, which is shown later. Formally,  $\Pi_{\mathbf{R}}$  corresponds to the Möbius inverse of  $I_{\min}$  [8]. And from this relationship, it is clear that  $\Pi_{\mathbf{R}}$  can be calculated recursively as

$$\Pi_{\mathbf{R}}(S; \alpha) = I_{\min}(S; \alpha) - \sum_{\beta \prec \alpha} \Pi_{\mathbf{R}}(S; \beta) \quad (28)$$

Put into words,  $\Pi_{\mathbf{R}}(S; \alpha)$  quantifies the information provided redundantly by the sources of  $\alpha$  that is not provided by any simpler collection of sources (i.e., any  $\beta$  lower than  $\alpha$  on the redundancy lattice shown above). **Note** that  $\Pi_{\mathbf{R}}$  is non-negative, which proof can be seen in [7].

Above all, we've formed a basic idea of the decomposition of partial information. We'll proceed to formulate how to measure these notions with set measure in the next chapter.

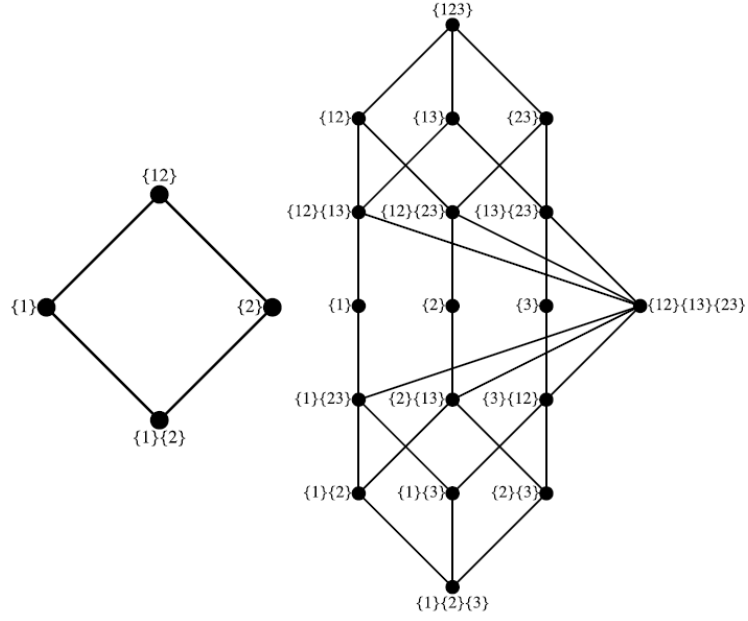


Figure 5: Redundancy lattice for (A) 3 and (B) 4 variables.

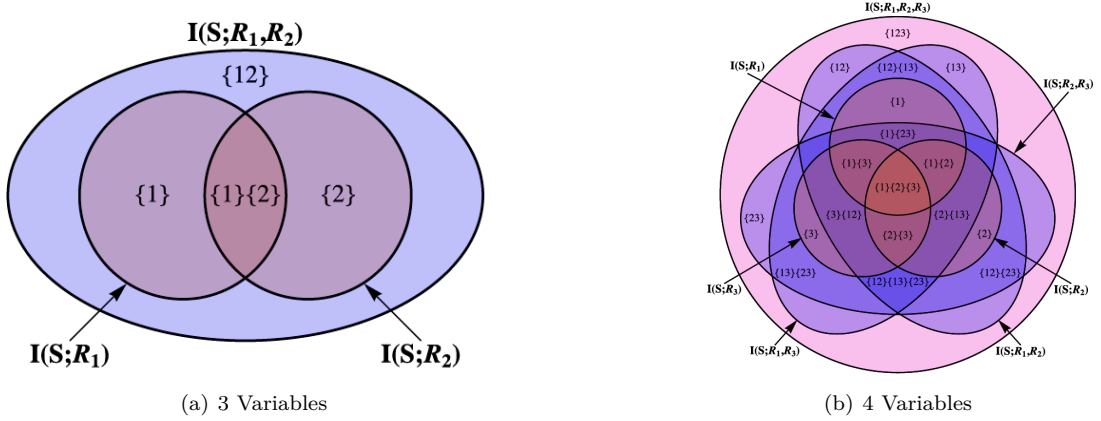


Figure 6: Partial Information Diagrams



## 4 PI-Measure

We next show how redundancy measure can be formulated in a set-theoretic way. In this chapter, we develop a theory which establishes a one-to-one correspondence between set theory and partial information measures, as well as their decomposed values, namely  $I_{\min}$  and  $\Pi_{\mathbf{R}}$ . Unlike the scenario where relationship between entropies and mutual information is formed, the correspondence here involves more atomic sets and more complicated set properties.

### 4.1 Formulation

In this section, using the notion of measure theory introduced in Chapter 2, we will give a formal definition for the correspondence. To fix ideas, we first formulate in this section the one-to-one correspondence between redundancy measures and set theory for a three dimensional random source vector.

**Field Generation** For random variables  $A_1, A_2$  and  $A_3$  in  $\mathbf{A}$ , let  $\tilde{A}_1, \tilde{A}_2$  and  $\tilde{A}_3$  be the sets corresponding to the information they provide to  $S$  respectively, which is  $I(S; A_i)$ , or equivalently saying,  $I_{\min}(S; \{A_i\})$ . Note that the field generated by these three sets is not enough in representing the general picture. With a maximal field size of  $2^n$ , it fails to cover all possible combinations of information redundancy measures, since our partial information decomposition is based on the second-order power set of all sources.

As a result, a natural solution for our formulation is to take the first-order power set of sources as the atom field. In addition to  $\tilde{A}_1, \tilde{A}_2$  and  $\tilde{A}_3$ , now we have to introduce  $\tilde{A}_{12}, \tilde{A}_{23}, \tilde{A}_{13}$  and  $\tilde{A}_{123}$ .

Formally, let  $A_1, \dots, A_k$  be nonempty and potentially dependent dimensions of sources  $\mathbf{A}$ . since the redundancy of a particular subset can be absorbed by its super-set, the field  $\mathcal{F}$  is defined as any collections of sets which can be obtained by any sequence of usual set operations on *basic sets*  $\Gamma$ , where

$$\Gamma = \{\beta \in \mathcal{P}_1(\mathbf{A}) : \forall A_i, A_j \in \beta, A_i \not\subseteq A_j\} \quad (29)$$

For every random variable collection in the  $\Gamma$ , its self-redundancy is defined as  $I_{\min}(S; \{A_i, \dots, A_j\})$ . We use  $\{\tilde{A}_i, \dots, \tilde{A}_j\}$ , or abbreviation  $\tilde{A}_{i\dots j}$  to represent the basic set its self redundancy corresponds to. Based on observations in Figure 6, the set measure we've defined has the following correspondence with redundancy measure through operation of intersection.

$$\mu^*(\tilde{A}_{i\dots j} \cap \dots \cap \tilde{A}_{m\dots n}) = I_{\min}(S; \{A_i, \dots, A_j\}, \dots, \{A_m, \dots, A_n\}) \quad (30)$$

**Symbolic Substitution** The equation above fails to cover other kinds of set operations like difference and union. In fact, due to the complexity of partial information, we can't write the direct PI correspondence of these operations.

Luckily, Chapter 3.3 has provided us with a useful tool, partial information  $\Pi_{\mathbf{R}}$ . Recall that partial information is the basic, non-intersecting building block of any redundancy measures, including their arbitrary sums of differences. With partial information, we can formally write out the symbol correspondence between sets and redundancy measures. The substitution strategy is listed as follows.

1. For every basic sets  $\tilde{A}_{i\dots j}$ ,  $\mu^*(\tilde{A}_{i\dots j}) = I_{\min}(S; \{A_i, \dots, A_j\})$ .
2. For any two sets  $\alpha$  in the field,  $\mu^*(\alpha^c) = \sum_{\alpha \prec \gamma} \Pi_{\mathbf{R}}(S; \gamma)$
3. For any two sets  $\alpha$  and  $\beta$  in the field,  $\mu^*(\alpha \cap \beta) = \sum_{\gamma \preceq \alpha \text{ and } \gamma \preceq \beta} \Pi_{\mathbf{R}}(S; \gamma)$
4. For any two sets  $\alpha$  and  $\beta$  in the field,  $\mu^*(\alpha \cup \beta) = \sum_{\gamma \preceq \alpha \text{ or } \gamma \preceq \beta} \Pi_{\mathbf{R}}(S; \gamma)$
5. For any two sets  $\alpha$  and  $\beta$  in the field,  $\mu^*(\alpha - \beta) = \sum_{\gamma \preceq \alpha \text{ and } \beta \prec \gamma} \Pi_{\mathbf{R}}(S; \gamma)$

By the definition of fields, every set in the diagram can be represented as a  $\cap$  or  $\cup$  combination of basic sets or its complement. With the strategy above, we have formally defined the *PI-Measure* in terms of measure theory.

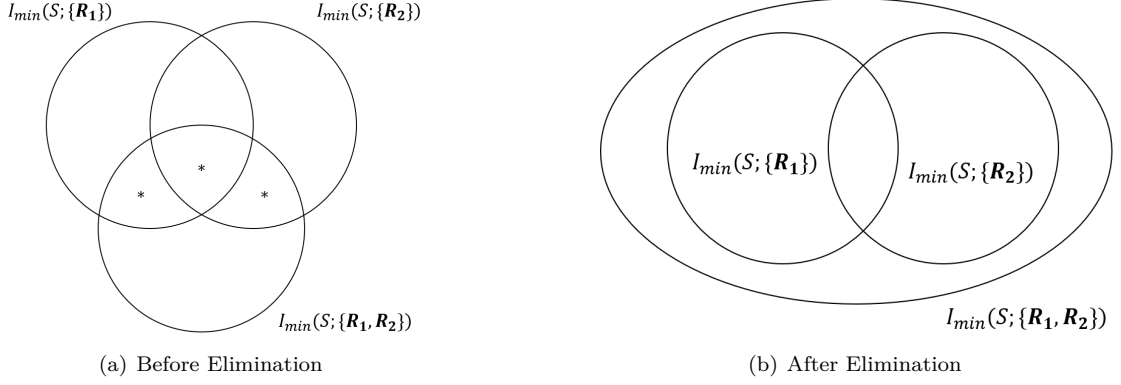


Figure 7: Field Elimination for 2 Source Variables

## 4.2 Simplification

There are two problems with our definition. One is that the basic sets, sets that generate fields are growing too fast as the number of random variable increases. And the other is that the correspondence we've defined for arbitrary sets is too complex to be of practical use. In this section, we will try to eliminate redundant fields in order to address the first problem and find alternative, useful expressions of fields by means of set theory.

**Eliminate Redundant Fields** An important usage of set-theory in information theory is to use Venn diagram to help derive formulas. Now, with the extra power-sets of sources as basic sets, it is impractical to plot redundancy information sets out as we did with Shannon information measures. However, the basic sets here are not independent with each other. For example,  $I(S; R_1, R_2)$  clearly has no complement intersecting with  $I(S; R_1)$  and  $I(S; R_2)$ . Hence, as what we did with Markov Chain in Chapter 2.3, we can also derive some useful formulas to eliminate non-existing fields, in order to simplify the diagram.<sup>1</sup>

To fix ideas, we take the multivariate information of three variables  $I(S; A_1, A_2)$  as example, note that  $\{\tilde{A}_1\} \subseteq \{\tilde{A}_1, \tilde{A}_2\}$ . It follows that

$$\begin{aligned} \mu^* \left( \left\{ \tilde{A}_1 \right\} \cap \left\{ \tilde{A}_1, \tilde{A}_2 \right\}^C \right) &= \mu^* \left( \left\{ \tilde{A}_1 \right\} \right) - \mu^* \left( \left\{ \tilde{A}_1 \right\} \cap \left\{ \tilde{A}_1, \tilde{A}_2 \right\} \right) \\ &= I_{\min}(S; \{\mathbf{A}_1\}) - I_{\min}(S; \{\mathbf{A}_1\}, \{\mathbf{A}_1, \mathbf{A}_2\}) = 0 \end{aligned} \quad (31)$$

Similarly, with the observation that  $\{\tilde{A}_2\} \subseteq \{\tilde{A}_1, \tilde{A}_2\}$ , we have  $\mu^* \left( \left\{ \tilde{A}_2 \right\} \cap \left\{ \tilde{A}_1, \tilde{A}_2 \right\}^C \right) = 0$ . As has been proved in 3.3 that all regions should have a non-negative measure, the sub-regions will also preserve a zero measure. Hence, three starred regions in Figure 7 will have a zero measure. To make our diagram more legible, we can erase these three regions out.

As the number of variables increases, the complexity of the combination also grows exponentially. It becomes nontrivial to identify all the subset relations, especially when higher-order fields are generated. Here is a formulated rule to follow.

**Theorem 8** (Field Elimination). *For any two basic sets  $\tilde{A}_{i\dots j}, \tilde{A}_{i'\dots j'} \in \Gamma$ , if  $\tilde{A}_{i\dots j} \subseteq \tilde{A}_{i'\dots j'}$ , then the measure of any set intersection with  $\tilde{A}_{i\dots j} \cap \tilde{A}_{i'\dots j'}^C$  is zero.*

*Proof.* The proof that  $\mu^* \left( \left\{ \tilde{A}_{i\dots j} \right\} \cap \left\{ \tilde{A}_{i'\dots j'} \right\}^C \right) = 0$  follows a similar pattern as Equation 31. For any set  $\eta$  that intersects with  $\left\{ \tilde{A}_{i\dots j} \right\} \cap \left\{ \tilde{A}_{i'\dots j'} \right\}^C$ , by the non-negativity of partial information measure, we have

$$0 \leq \mu^* \left( \eta \cap \left\{ \tilde{A}_{i\dots j} \right\} \cap \left\{ \tilde{A}_{i'\dots j'} \right\}^C \right) \leq \mu^* \left( \left\{ \tilde{A}_{i\dots j} \right\} \cap \left\{ \tilde{A}_{i'\dots j'} \right\}^C \right) = 0 \quad (32)$$

<sup>1</sup>We should not confuse empty sets with sets that have a measure of zero. In other words, the reason why we don't plot these sets out is because their measures are zero, not that these sets don't 'exist'.

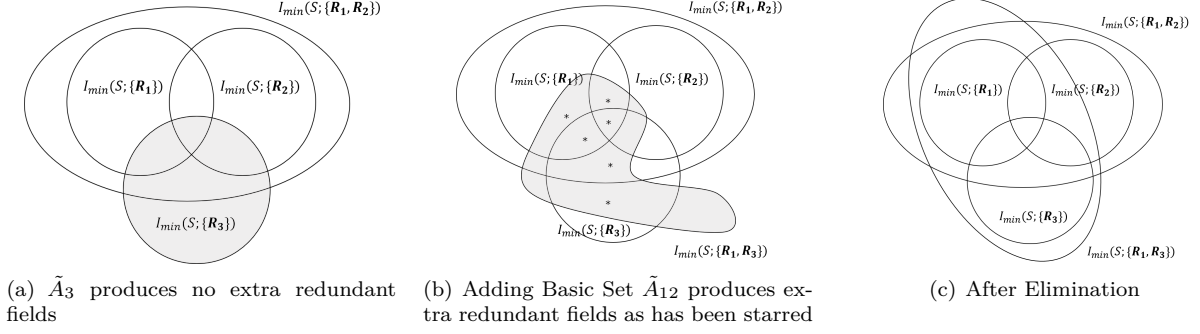


Figure 8: Field Elimination Strategy for 3 Source Variables

□

The above theorem also implies a strategy to plot a pretty partial information diagram. Starting with singleton basic sets, every time we draw a new basic set, check the subset relation with the existing sets and eliminate the intersecting fields out. A construction of partial information diagram for four random variables is shown in Figure 8. With such strategy, we can make clear how the partial information diagram in Figure 6 is step-by-step formulated with a set theoretic view.

**Alternative Field Expressions** Note that the partial information  $\Pi_{\mathbf{R}}$  is actually a derived concept. In common cases, we are more interested in the redundancy measure, which has more practical implications. We try to derive the relation between arbitrary field and linear combinations of redundancy measure from a pure set-theoretic view.

Note that the measure of *atoms* in  $\mathcal{F}_n$  are a perfect matching to the partial information  $\Pi_{\mathbf{R}}$ . The reason is that by Definition 27,  $\Pi_{\mathbf{R}}(S; \alpha)$  only provides the unique information from  $\alpha$ . There is no possibility that the sets corresponding to a single partial information can be deposited to smaller sets by intersection, which coincides with the property of set atoms.

With the observation above, we can narrow our scope into atoms, since the measure of arbitrary set can be written into a combination of atoms. Because all atoms can be represented as sets of

$$\cap_{i \in \mathcal{N}_n} \tilde{Y}_i (\tilde{Y}_i = \tilde{A}_i \text{ or } \tilde{A}_i^c), \text{ where the index set } \mathcal{N}_n = \mathcal{P}_1(\{1, 2, \dots, n\}) \quad (33)$$

we only need to study two operations, namely complement and intersection, on basic sets  $Y_{i \in \mathcal{N}_n}$ . The mapping of such operations to redundancy measures can be listed as follows.

$$\mu^* \left( \tilde{A}_{N_1} \cap \dots \cap \tilde{A}_{N_k} \cap \tilde{A}_{M_1}^C \cap \dots \cap \tilde{A}_{M_j}^C \right) \quad (34)$$

$$= \mu^* \left( \tilde{A}_{N_1} \cap \dots \cap \tilde{A}_{N_k} \right) - \sum_{i=1}^k \mu^* \left( \tilde{A}_{N_1} \cap \dots \cap \tilde{A}_{N_k} \cap \tilde{A}_{M_i} \cap \dots \cap \tilde{A}_{M_j} \right) \quad (35)$$

$$\text{where } \{N_1, \dots, N_k, M_1, \dots, M_j\} = \mathcal{N}_n \quad (36)$$

The formula above is not hard to obtain from a pure set-theoretic perspective, now note that in Equation 35, all the sets being measured are intersections of basic sets. By applying Equation 30, we can write them out in the form of redundancy measure  $I_{\min}$ . By now, we have finished proving that any atom can be written as a  $\pm 1$  combination of some redundancy measures.

Therefore, we can conclude that any field can be written as a linear combination of redundancy measures. An application of such expression is to calculate partial information in a top-down way. In the next chapter we will apply such strategy to show how our method can eliminate the need to calculate lots of partial information in the traditional bottom-up strategy.

Note that if we change the equality into  $\subseteq$  in Equation 36, the formula will still hold, i.e. we can always write out the intersection of any basic sets or their complements in the form of Equation 35, which may help us reduce the complexity of computing redundancy a step further.

### 4.3 Completeness

In order to validate the effectiveness of the PI-measure correspondence, we will close our formulation section with a conclusion that the measure of any field can be completely obtained from the measures of basic sets.

**Theorem 9.** *Let*

$$\mathcal{B} = \left\{ \tilde{X}_G : G \text{ is a nonempty subset of } \mathcal{N}_n \right\} \quad (37)$$

*Then a measure  $\mu^*$  on  $\mathcal{F}_n$  is completely specified by  $\{\mu^*(B), B \in \mathcal{B}\}$ , which can be any set of real numbers.*

*Proof.* The number of elements in  $\mathcal{B}$  is equal to the number of nonempty subsets of  $\mathcal{N}_n$ , which is  $2^{2^n-1} - 1$ . Thus  $|\mathcal{A}| = |\mathcal{B}| = 2^{2^n-1} - 1$ . Let  $k = 2^{2^n-1} - 1$ . Let  $\mathbf{u}$  be a column  $k$ -vector of  $\mu(A), A \in \mathcal{A}$ , and  $\mathbf{h}$  be a column  $k$ -vector of  $\mu(B), B \in \mathcal{B}$ . since all the sets in  $\mathcal{B}$  can be expressed uniquely as the union of some nonempty atoms in  $\mathcal{A}$ , by the set-additivity of  $\mu$ , for each  $B \in \mathcal{B}$ ,  $\mu(B)$  can be expressed uniquely as the sum of some components of  $\mathbf{u}$ . Thus

$$\mathbf{h} = C_n \mathbf{u} \quad (38)$$

where  $C_n$  is a unique  $k \times k$  matrix. On the other hand, it has been shown that for each  $A \in \mathcal{A}$ ,  $\mu^*(A)$  can be expressed as a linear combination of  $\mu(B), B \in \mathcal{B}$  by applications, if necessary, of Equation 35.

However, the existence of the said expression does not imply its uniqueness. Nevertheless, we can write

$$\mathbf{u} = D_n \mathbf{h} \quad (39)$$

for some  $k \times k$  matrix  $D_n$ . Upon substituting Equation 38 into Equation 39, we obtain

$$\mathbf{u} = (D_n C_n) \mathbf{u} \quad (40)$$

which implies that  $D_n$  is the inverse of  $C_n$  as Equation 40 holds regardless of the choice of  $\mu$ . since  $C_n$  is unique, so is  $D_n$ . Therefore,  $\mu^*(A), A \in \mathcal{A}$  are uniquely determined once  $\mu^*(B), B \in \mathcal{B}$  are specified. Hence, a measure  $\mu^*$  on  $\mathcal{F}_n$  is completely specified by  $\{\mu(B), B \in \mathcal{B}\}$ , which can be any set of real numbers. The theorem is proved.  $\square$

## 5 Application

In Chapter 4, we introduce the measure of partial information, which successfully captures the unique information provided by every possible combination of sources. Then we know the total information provided by a source can be divided into several partial information contributions, because the mutual information becomes a sum of these partial information terms.

Here we introduce the application of our theories. We try to calculate the value of any  $\Pi_R$  in two ways. On one hand, we can use the original definition of  $\Pi_R$  to calculate, by applying Equation 41. On the other hand, we can calculate  $\Pi_R$  by converting it to the linear combination of  $I_{\min}$  with our strategy proposed above.

**Three Variables** First consider the simplest case of a system with three variables:  $S, R_1$  and  $R_2$ , we want to calculate the value of  $\Pi_R(S; \{1\})$ .

With the original definition, we can apply  $\beta = \{\{1\}\{2\}\}$  to the following Equation

$$\Pi_R(S; \alpha) = I_{\min}(S; \alpha) - \sum_s p(s) \max_{\beta \in \alpha^-} \min_{\mathbf{B} \in \beta} I(S = s; \mathbf{B}) \quad (41)$$

to get

$$\Pi_R(S; \{1\}) = I_{\min}(S; \{1\}) - I_{\min}(S; \{1\}\{2\}) \quad (42)$$

Use our theorem, based on the observations in Figure 6(a), we can express the field using Equation 30:

$$\begin{aligned} \mu^* \left( \left( \tilde{R}_1 \right) \cap \left( \tilde{R}_2 \right)^C \cap \left( \tilde{R}_1, \tilde{R}_2 \right) \right) &= \mu^* \left( \left( \tilde{R}_1 \right) \cap \left( \tilde{R}_1, \tilde{R}_2 \right) \right) - \mu^* \left( \left( \tilde{R}_1 \right) \cap \left( \tilde{R}_2 \right) \cap \left( \tilde{R}_1, \tilde{R}_2 \right) \right) \\ &= I_{\min}(S; \{1\}, \{1, 2\}) - I_{\min}(S; \{1\}, \{2\}, \{1, 2\}) \\ &= I_{\min}(S; \{1\}) - I_{\min}(S; \{1\}, \{2\}) \end{aligned} \quad (43)$$

where the last equality holds because  $\{\tilde{R}_1\} \subseteq \{\tilde{R}_1, \tilde{R}_2\}$ .

We can show the equivalence of the two methods with a simple example. In Figure9, the black tiles represent equally-probable outcomes and the white tiles are zero-probability outcomes. Using the formula below

$$\begin{aligned}
I(S; \mathbf{A}) &= \sum_s p(s) I(S = s; \mathbf{A}) \\
I(S = s; \mathbf{A}) &= \sum_{\mathbf{a}} p(\mathbf{a}|s) \left[ \log \frac{1}{p(s)} - \log \frac{1}{p(s|\mathbf{a})} \right] \\
I_{\min}(S; \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k\}) &= \sum_s p(s) \min_{\mathbf{A}_i} I(S = s; \mathbf{A}_i)
\end{aligned} \tag{44}$$

we can calculate that  $I(S; \{1\}) = I(S; \{2\}) = -\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3}$ , and  $I_{\min}(S; \{1\}, \{2\}) = \log 3 - \log 2$ . Finally, we have  $\Pi_{\mathbf{R}}(S; \{1\}) = -\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} - \log 3 + \log 2 = \frac{1}{3}$ .

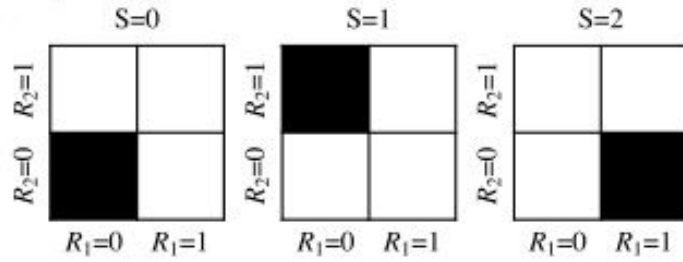


Figure 9: Probability distributions for  $S \in \{0, 1, 2\}$  and  $R_1, R_2 \in \{0, 1\}$ .

**Four Variables** Now we consider a more complicated system with four variables, and we want to calculate the value of  $\Pi_{\mathbf{R}}(S; \{2\}, \{13\})$ .

On one hand, we can apply Equation 41 to solve the problem:

$$\Pi_{\mathbf{R}}(S; \{2\}, \{13\}) = I_{\min}(S; \{2\}, \{13\}) - \sum_s p(s) \max_{\beta \in \alpha^-} \min_{\mathbf{B} \in \beta} I(S = s; \mathbf{B}) \tag{45}$$

where  $\beta \in \{\{\{1\}\{2\}\}, \{\{2\}\{3\}\}\}$ . Then we just need to expand the formula to get the result.

On the other hand, based on the observations in Figure6(b), we can express the field using Equation 30:

$$\begin{aligned}
& \mu^* \left( \left\{ \tilde{R}_1 \right\}^C \cap \left\{ \tilde{R}_2 \right\} \cap \left\{ \tilde{R}_3 \right\}^C \cap \left\{ \tilde{R}_{12} \right\} \cap \left\{ \tilde{R}_{13} \right\} \cap \left\{ \tilde{R}_{23} \right\} \cap \left\{ \tilde{R}_{123} \right\} \right) \\
&= \mu^* \left( \left\{ \tilde{R}_2 \right\} \cap \left\{ \tilde{R}_{12} \right\} \cap \left\{ \tilde{R}_{13} \right\} \cap \left\{ \tilde{R}_{23} \right\} \cap \left\{ \tilde{R}_{123} \right\} \right) - \mu^* \left( \left\{ \tilde{R}_1 \right\} \cap \left\{ \tilde{R}_2 \right\} \cap \left\{ \tilde{R}_{12} \right\} \cap \left\{ \tilde{R}_{13} \right\} \cap \left\{ \tilde{R}_{23} \right\} \cap \left\{ \tilde{R}_{123} \right\} \right) \\
&\quad - \mu^* \left( \left\{ \tilde{R}_2 \right\} \cap \left\{ \tilde{R}_3 \right\} \cap \left\{ \tilde{R}_{12} \right\} \cap \left\{ \tilde{R}_{13} \right\} \cap \left\{ \tilde{R}_{23} \right\} \cap \left\{ \tilde{R}_{123} \right\} \right) \\
&\quad + \mu^* \left( \left\{ \tilde{R}_1 \right\} \cap \left\{ \tilde{R}_2 \right\} \cap \left\{ \tilde{R}_3 \right\} \cap \left\{ \tilde{R}_{12} \right\} \cap \left\{ \tilde{R}_{13} \right\} \cap \left\{ \tilde{R}_{23} \right\} \cap \left\{ \tilde{R}_{123} \right\} \right) \\
&= I_{\min}(S; \{2\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}) - I_{\min}(S; \{1\}, \{2\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}) \\
&\quad - I_{\min}(S; \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}) + I_{\min}(S; \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}) \\
&= I_{\min}(S; \{2\}, \{1, 3\}) - I_{\min}(S; \{1\}, \{2\}) - I_{\min}(S; \{2\}, \{3\}) + I_{\min}(S; \{1\}, \{2\}, \{3\})
\end{aligned} \tag{46}$$

By some set operations, it can be translated into the linear combination of  $I_{\min}$ . The last equality holds because the former can be greatly simplified. For example,  $\{\tilde{R}_2\} \subseteq \{\tilde{R}_1, \tilde{R}_2\}, \{\tilde{R}_2, \tilde{R}_3\} \dots$  Now we just need to calculate the value of each  $I_{\min}$ .

Now we use another simple example to illustrate the equivalence between the two expressions. Based on the probability distribution in Figure 9, we introduce another variable  $R_3$ . Consider the equally-probable cases of  $S$  as follows:

$$\begin{aligned} S = 0 &\iff R_1 = 0, R_2 = 0, R_3 = 0 \\ S = 1 &\iff R_1 = 0, R_2 = 1, R_3 = 1 \\ S = 2 &\iff R_1 = 1, R_2 = 0, R_3 = 0 \end{aligned}$$

On one hand, by applying Equation 45, we have

$$I_{\min}(S; \{2\}, \{13\}) = \log 3 - \frac{2}{3} \log 2 \quad (47)$$

$$\sum_s p(s) \max_{\beta \in \alpha^-} \min_{\mathbf{B} \in \beta} I(S = s; \mathbf{B}) = \log 3 - \frac{2}{3} \log 2 \quad (48)$$

So  $\Pi_R(S; \{2\}, \{13\})=0$ .

On the other hand, by calculating by the linear combination of  $I_{\min}$ , we can also have the same outcome. In other words,

$$I_{\min}(S; \{1\}, \{2\}) - I_{\min}(S; \{2\}, \{3\}) + I_{\min}(S; \{1\}, \{2\}, \{3\}) = \sum_s p(s) \max_{\beta \in \alpha^-} \min_{\mathbf{B} \in \beta} I(S = s; \mathbf{B}) \quad (49)$$

## 6 Discussion

In this project, we first revisited I-Measure, a correspondence between Shannon's information and set theory. Then we introduced redundancy measure, together with its decomposed form, partial information, a non-negative measure in describing the redundancy a multivariate collection of sources provide for a particular random variable.

Inspired by I-Measure, we formally defined "PI-Measure" in the scope of set theory to formally describe how partial information diagrams are formed. We formulate the correspondence rules between PI-Measure and gave out its correspondence with partial information and redundancy measure. Then we made some simplifications to eliminate redundant fields. We've also found a strategy to find alternative expressions on partial information from a pure set-theoretic perspective. Finally, we calculated the partial information for some simple examples to compare our proposed strategy with the existing one.

We hope that our formulation can help clarify and substantiate how partial information diagrams are generated and can be used.

## References

- [1] Claude E Shannon. "A mathematical theory of communication". In: *Bell system technical journal* 27.3 (1948), pp. 379–423 (cit. on p. 1).
- [2] Raymond W Yeung. *Information theory and network coding*. Springer Science & Business Media, 2008 (cit. on p. 1).
- [3] Raymond W Yeung. *A first course in information theory*. Springer Science & Business Media, 2012 (cit. on p. 1).
- [4] Spartak T Belyaev et al. *The Many Body Problem*. 1959 (cit. on p. 1).
- [5] P. Dayan and L F Abbott. "Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems". In: (2001) (cit. on p. 1).
- [6] Fred Rieke et al. "Spikes, Exploring the Neural Code". In: (1996) (cit. on p. 1).
- [7] Paul L. Williams and Randall D. Beer. "Nonnegative Decomposition of Multivariate Information". In: (2010) (cit. on pp. 1, 7).
- [8] RP Stanley. "Enumerative combinatorics, vol. 1 Cambridge University Press Cambridge". In: *England zbMATH* (1997) (cit. on p. 7).