

CHAPTER 1

Venn Diagrams and Information Theory

“The reader will find many figures in this work.¹”

§1.1 Venn Diagrams and Information Theory

Venn diagrams used to understand the information theoretic relationships among random variables, or *I-diagrams*, are a crucial member of any information theorist’s tool belt (see Fig. 1.2).

Given a set of random variables, there is a one-to-one mapping from the set of information quantities—entropies, mutual informations, conditional mutual informations—to an I-diagram.

Given that this map is one-to-one, how can I claim that the I-diagram is crucial? Is this not just another redundant representation to learn? The answer is most definitely: No.

As is the general theme of this thesis and the surrounding body of work, *natural* representations are given special status and attention. The I-diagram is not claimed to be natural in the same sense that the ϵ -machine is. However, the association of random variables with geometric bodies, and their relationships with geometric intersection and occlusion, will *prove itself* to be natural for humans.

Why might this be the case? I claim that this is due to our visual hardware/firmware. We very naturally perceive connected homogeneous regions as objects. We also are skilled at understanding occlusion and how occlusions change as objects move. Then again, for some reason it seemed to take man a rather long time to figure out what the lunar eclipse was.

I submit that the student of information theory *knows* I-diagrams whether or not they have seen one.² Here we seek only to make this knowledge more explicit and to develop a little of the I-diagram calculus.

¹“The reader will find no figures in this work” - J. L. Lagrange, *Mécanique Analytique*, 1888.

²“Before hearing Monge, I did not know that I knew descriptive geometry.” - J. L. Lagrange, said after a lecture by Gaspard Monge. Monge was the inventor of descriptive geometry. This highly intuitive method is now integral to design and engineering.

§1.2 How to read an I-diagram

Each random variable corresponds to one ‘smooth’ object in the diagram. This is not a requirement, but makes visual inspection easier. When possible, we will use circles or ellipses. Certain circumstances will encourage slightly more exotic smooth shapes.

The area of a smooth shape corresponds to the entropy of that random variable, $H[X]$ (see Fig. 1.1). This is the total uncertainty in the variable X . We will loosely refer to these shapes as both random variables and their entropies. The meaning should be clear from context.

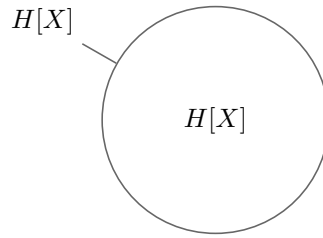


Figure 1.1: The simplest I-diagram - one random variable, X .

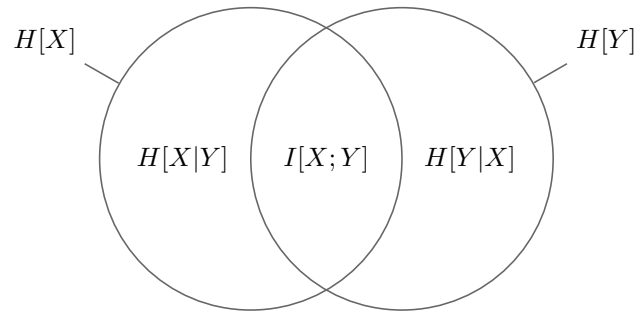


Figure 1.2: I-diagram for two random variables, X and Y .

In an I-diagram involving two random variables, the area of intersection of two random variable shapes, X , Y , corresponds to the mutual information between those random variables, $I[X; Y]$ (see Fig. 1.3). This is the information that the two random variables share.

The area of random variable X that is unoccluded by random variable Y corresponds to the conditional entropy, $H[X|Y]$. This is the uncertainty in X that remains when Y is known.

Now consider a three variable I-diagram. What is the uncertainty in X given *both* Y and Z ? This is notated $H[X|YZ]$ as in Fig. 1.5.

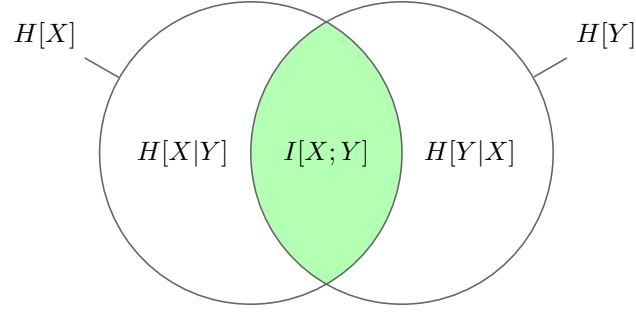


Figure 1.3: The mutual information between X and Y is highlighted.

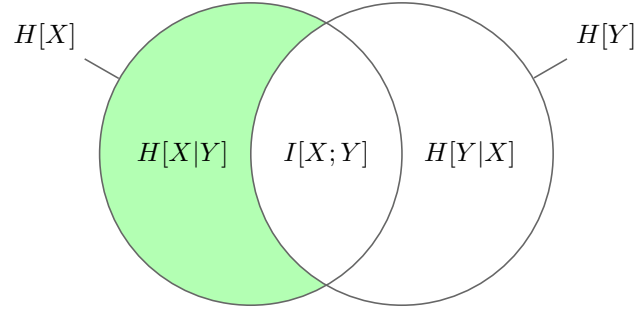


Figure 1.4: The conditional entropy of X given Y is highlighted.

We can equivalently think of merging Y and Z to form a random variable, W , with a larger event space. In this light, our quantity of interest would look like $H[X|W]$. We can think of this graphically quite naturally by considering W as the union of Y and Z (see Fig. 1.6).

Now considering a three variable I-diagram, the area of intersection between X and Y that is unoccluded by Z is the conditional mutual information, $I[X; Y|Z]$ (see Fig. 1.7). This is the information shared by X and Y when Z is known. Adding knowledge about Z can either increase or decrease the (conditioned) mutual information.

From these diagrams, e.g. Fig. 1.2, we can visually prove such identities as:

$$H[XY] = H[X|Y] + I[X; Y] + H[Y|X]$$

$$I[X; Y] = H[X] - H[X|Y]$$

§1.2.1 Stratification of a Composite Variable

Just as more than one random variable can be considered as a composite variable, as in Fig. 1.6, a composite variable can be stratified in a very intuitive way. Consider the set of random variables $\{X_0, X_1, X_2, X_3, X_4\}$. The composite random variable X_0^5 can be stratified by the set $\{X_0^1, X_0^2, X_0^3, X_0^4, X_0^5\}$. These results are generic for any set of ‘stratifying’ variables.

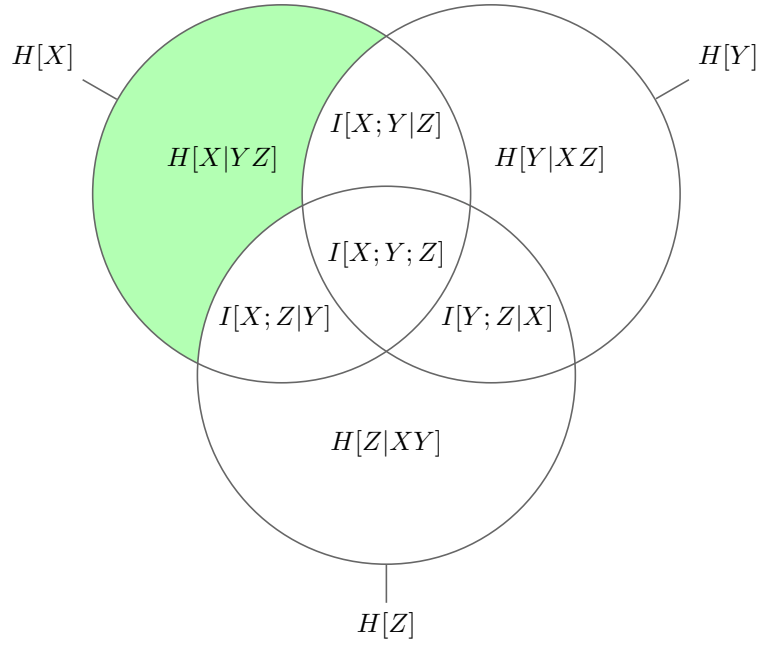


Figure 1.5: The conditional entropy of X given Y and Z is highlighted.

Definition. Given a set $\mathbf{X} = \{X_1, \dots, X_N\}$ for some integer, N , a set of subsets of \mathbf{X} , \mathbf{Y} , is called stratifying if a total ordering on \mathbf{Y} is induced by the relation \subset .

For any such \mathbf{X} , we have the stratifying set $\mathbf{Y} = \{\{X_1\}, \{X_1, X_2\}, \dots, \{X_1, X_2, \dots, X_N\}\}$.

We can stratify the past random variables of a process as in Fig. 1.8.

That this is true follows straightforwardly from the duality between information measures and set operations demonstrated by [cite Yeung](#).

Additionally, the I-diagram involving a stratified variable and one additional variable has the form in Fig. ???. To see that this is not a trivial statement, consider the following scenario. Let random variables X and Y each be fair coins. Then let Z be the exclusive-or (XOR) of X and Y . If we were to begin with the I-diagram for just X and Y , we would have the following picture. [graphic here](#)

Then naively ‘adding’ Z to the I-diagram yields the following picture. [graphic here](#)

This fails to capture the true information theoretic structure. Specifically, it does not capture that $I[X; Y|Z] = 1$ or that $I[X; Y; Z] = -1$. The correct picture does not allow for regions to be eliminated before the addition of all variables to the diagram. This is why it is non-trivial to ‘naturally’ add Z to the stratified variable I-diagram.

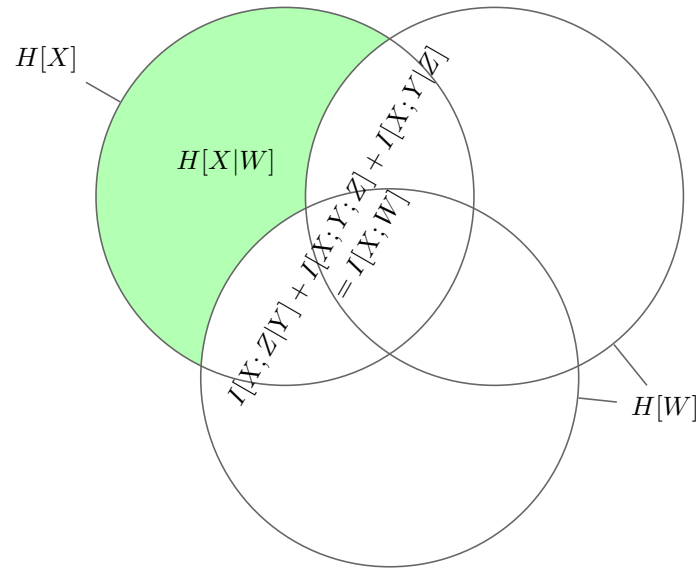


Figure 1.6: The mutual information of X and joint variable W is highlighted.

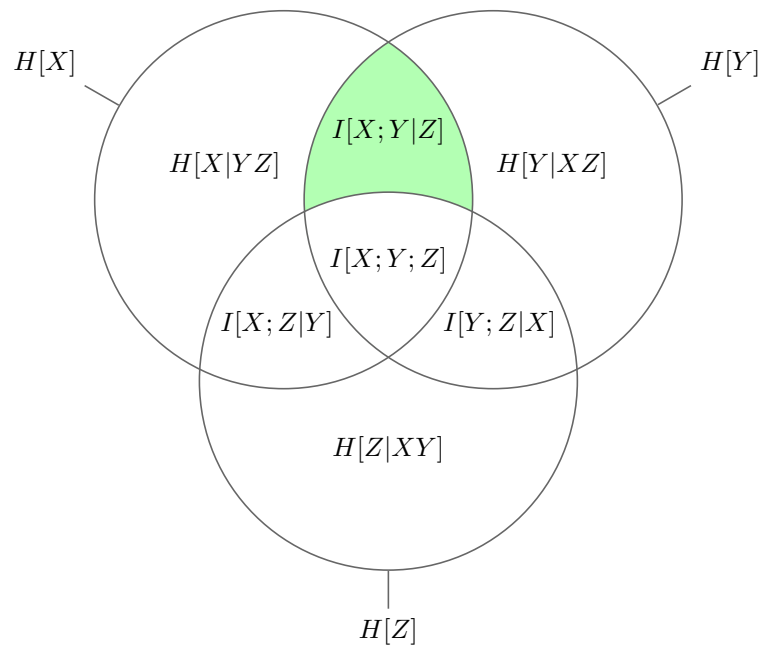


Figure 1.7: The conditional mutual entropy of X and Y given Z is highlighted.

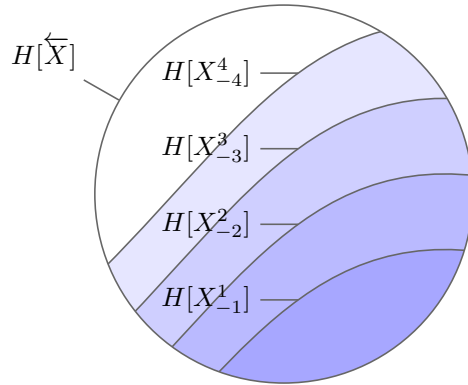


Figure 1.8: The standard stratification of the conglomerate random variable \overleftarrow{X} .