

Information Theory

Govind G S

June 26 2022

1 Introduction

Information theory is the scientific study of the quantification, storage, and communication of digital information.[1] The field was fundamentally established by the works of [Harry Nyquist](#) and [Ralph Hartley](#), in the 1920s, and [Claude Shannon](#) in the 1940s. The field is at the intersection of probability theory, statistics, computer science, statistical mechanics, information engineering, and electrical engineering.

2 Entropy

In information theory, the entropy of a random variable is the average level of "information", "surprise", or "uncertainty" inherent to the variable's possible outcomes.

2.1 Equation

Given a discrete random variable X , which takes values in the alphabet \mathcal{X} and is distributed according to $p : \mathcal{X} \rightarrow [0, 1]$:

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

The symbols used and their meaning are referenced in Table ??

2.2 Base of log

The choice of base for log , the logarithm, varies for different applications. Base 2 gives the unit of bits (or "shannons"), while base e gives "natural units" nat, and base 10 gives units of "dits", "bans", or "hartleys". An equivalent definition of entropy is the expected value of the self-information of a variable.

3 The Core Idea

The core idea of information theory is that the "informational value" of a communicated message depends on the degree to which the content of the message is surprising. If a highly likely event occurs, the message carries very little information. On the other hand, if a highly unlikely event occurs, the message is much more informative. For instance, the knowledge that some particular number will not

Symbol	Meaning
H	Entropy
X	Discrete random variable
x_i	Possible values of X
P	Probability Mass Function of each x_i

Table 1: Symbols used in the Equation26

be the winning number of a lottery provides very little information, because any particular chosen number will almost certainly not win. However, knowledge that a particular number will win a lottery has high informational value because it communicates the outcome of a very low probability event

4 Properties

The Shannon entropy satisfies the following properties, for some of which it is useful to interpret entropy as the expected amount of information learned (or uncertainty eliminated) by revealing the value of a random variable X :

1. Adding or removing an event with probability zero does not contribute to the entropy:
$$H_{n+1}(p_1, \dots, p_n, 0) = H_n(p_1, \dots, p_n).$$
2. If X and Y are two independent random variables, then knowing the value of Y doesn't influence our knowledge of the value of X (since the two don't influence each other by independence):
$$H(X|Y) = H(X).$$
3. The entropy of two simultaneous events is no more than the sum of the entropies of each individual event i.e., $H(X, Y) \leq H(X) + H(Y)$, with equality if and only if the two events are independent.
4. The entropy or the amount of information revealed by evaluating (X, Y) (that is, evaluating X and Y simultaneously) is equal to the information revealed by conducting two consecutive experiments: first evaluating the value of Y , then revealing the value of X given that you know the value of Y . This may be written as: $H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.