

Shannon's metric of "entropy" of data

[Education](#), [Learning](#)



Shannon's metric of "Entropy" of data could be a foundational thought of Info theory. Here is an intuitive manner of understanding, basic cognitive process, and/or reconstructing Shannon's Entropy metric for data.

Conceptually, Info is thought of as being kept in or transmitted as variables which will have completely different values. A variable is thought of as a unit of storage which will have, at completely different times, one in all many completely different such values, following some method for taking over those values. Informally, we have a tendency to get data from a variable by staring at its worth, even as we have a tendency to get data from an email by reading its contents. Within the case of the variable, the data is concerning the method behind the variable.

The entropy of a variable is that the "amount of information" contained within the variable. This quantity is decided not simply by the quantity of various values the variable will have, even as the data in an email is quantified not simply by the quantity of words within the email or the various doable words within the language of the e-mail. Informally, the number of data in an email is proportional to the number of "surprise" its reading causes. As an example, if an email is just a repeat of an earlier email, then it's not informative in the least. On the opposite hand, if say the e-mail reveals the result of a cliff-hanger election, then it's extremely informative. Similarly, the data during a variable is tied to the number of surprise that worth of the variable causes once disclosed.

Shannon's entropy quantifies the number of data during a variable, so providing the inspiration for a theory around the notion of data.

Storage and transmission of data will intuitively be expected to be tied to the number of data concerned. as an example, data could also be concerning the result of a coin toss. This data is kept during a mathematician variable which will have the values zero or one. we will use the variable to represent the data of the coin toss, viz., whether or not the coin toss came up heads or not. In digital storage and transmission technology, this mathematician variable is drawn during a single “ bit”, the essential unit of digital data storage/transmission. However, this bit directly stores the worth of the variable, i. e. the data of the result of the coin toss. It doesn't compactly capture the data within the coin toss, e. g., whether or not the coin is biased or unbiased, and, if biased, how biased.

Whereas, Shannon's entropy metric quantifies, among alternative things, absolutely the minimum quantity of storage and transmission required for compactly capturing any data (as opposed to raw data), and in typical cases that quantity is a smaller amount than what's needed to store or transmit the raw data behind the {data}. Shannon's Entropy metric additionally suggests how of representing the data within the calculated fewer variety of bits.