Information Transfer

to be named

Mrenal Kanti Das Mridul

ABSTRACT:

Timeline: From March 03, 2025
To Month Date, Year

AUTHOR'S NOTE

Mrenal Kanti Das Mridul
March 03, 2025

CONTENTS

Chapter I: Glossary

Chapter I: Concept of Information

Chapter Contents

Chapter Overview

This section formalizes the concept of information in the terms of *probability* theory, *entropy*, and *information theory*. It provides mathematical definitions od *Shannon Entropy*, explores how a piece of information is measured in small bits, and atte,mpts to show the key results of information theory, such as *Kraft-McMillan Inequality*, and redundancy in encoding.

I.1: Conceptual Notion

"*It from bit. Otherwise put, every 'it' — every particle, every field of force, even the space-time continuum itself — derives its function, its meaning, its very existence entirely — even if in some contexts indirectly — from the apparatus-elicited answers to yes-or-no questions, binary choices, bits.*"

– John Archibald Wheeler, 1989.

In abstract terms, information can be understood as the *resolution of uncertainty*, i.e. it is what one entity gains when one learns something new. Such can be through any possible mediums like spoken language, written text, digital signals, stored media, or anything that can convey something meaningful which can then be process if acquired. However, information is not just only the content of a message, but it is also about its structure, probability, and efficiency in transmission.

One of the crucial insights of information theory is that *information is deeply connected to probability*, and thus, can be expressed mathematically. One understands information in terms of probability and uncertainty.

One considers a simple example of *coin toss*. A fair toss has two possible outcomes; heads or tails, each with a probability of 50%. When the coin is tossed, the outcome is uncertain, but the information gained from the outcome is very much significant, because there are two likely possibilities. In each case, the uncertainty is maximized, making heads or tails both likely to be the outcome. One also observes, that when there is heads, no tails, and vice versa. So, such an incident, or any random event, when the outcome is observed, one knows for sure that the remaining possibilities did not make it as the observed result.

For a biased case, meaning if the outcome of any of heads or tails are 75%, the other being 25%, making the first more predictable, and latter not; the uncertainty decreases. One sees that one gains less information because one outcome is more predictable than the other. It is a fine, yet very basic example of how the amount of information gained is related to the uncertainty of a random event.

I.2: Axiomatic Characterization

"*An axiom is a self-evident truth that requires no proof.*"

– Aristotle

Clade Shannon introduced the concept of *entropy* in his 1948 breakthrough paper, *"A Mathematical Theory of Communication"*, to quantify the information content associated with a random variable. What he sought was a function $H(p_1, p_2, p_3, \ldots, p_n)$ which measures the average information content of the random variable X satisfying the following properties:

Continuity: $H(X)$ should be a continuous function of p_i.

Maximum Entropy for Uniform Distribution: If all outcomes are to be equally likely, meaning $p_i = 1/n$, then entropy should be maximized.

Additivity (or Recursivity): If a system has multiple independent components, or parts, the total entropy should be the summation of the individual entropies. Mathematically, if X is split into two parts (Y, Z), the entropy has to satisfy:

$$H(X) = H(Y) + \textstyle\sum_j p_j H(Z|Y = y_i)$$

These properties, though seem arbitrary, are crucial for the characterization of Shannon Entropy, and ensures that the entropy serves as a consistent and reliable measure of uncertainty across various contexts.

Particularly the additivity is the foundation for information measuring. Additivity underpins the definition of mutual information, and conditional entropy, which are very important in quantifying information shared between variables and the uncertainty remaining in one variable given another. An alternative axiomatic characterization of Shannon Entropy involves properties like subadditivity, that states, For jointly distributed random variables X and Y;

$$H(X,Y) \leq H(X) + H(Y)$$

which indicates that the joint entropy of two variables are less than or equal to the sum of individual component's entropy. The equality holds when when X and Y are independent, which leads to the additivity;

$$H(X,Y) = H(X) + H(Y)$$

which reaffirms that the entropy of independent systems combines linearly.

The continuity implies that small changes in the distribution of probability of a random variable result in small changes in entropy. Formally, if P_X and P_Y are two probability distributions that are close in terms of total distance variation, then their entropies $H(X)$ and $H(Y)$ are also close. It ensures that such does not exhibit abrupt changes in minor alterations in the probability distribution, making it stable. Mathematically;

$$|H(X) - H(Y)| \leq K \cdot d_{TV}(P_X, P_Y)$$

where K is a positive arbitrary constant, and d_{TV} denotes total variation distance. In practical scenarios where a bit or piece of information may be subject to noise, or slight variations, continuity ensures that the entropy remains relatively unaffected, ensuring consistent performance in abstract systems which rely on the calculation of entropy.

As for the maximum entropy, it ensures that among all possible probability distributions over a set of outcomes, the uniform distribution maximizes entropy. Such means that the entropy is highest when all the outcomes are equally likely, which reflects maximum uncertainty. It aligns with the idea that, in the absence of some specific information, one should always assume all the outcomes that can be possible are probable, i.e. a maximum state of uncertainty. Such is noted as the *Principle of Insufficient Reason*, which is also the foundation of maximum entropy method in statistical mechanics and inference, where the least biased estimate possible is chosen based o the information, ensuring that no unwarranted assumptions are being made.

Formally; given a set of constraints, such as known expected values $\langle f_i(X) \rangle = c_i$ for functions f_i, the probability distribution $P(X)$ which maximizes the entropy is;

$$H(X) = -\textstyle\sum_{x \in X} P(x) \log P(x)$$

which is the most unbiased representation of the knowledge. In scenarios where least amount of information is available, it provides a systematic method for inferring probability distributions, leading to more reliable and generalizable models.

If one takes a random variable X which can take values $x_1, x_2, x_3, \ldots, x_n$ with corresponding probabilities $p(x_1), p(x_2), p(x_3), \ldots, p(x_n)$, where:

$$\sum_i^n p_i = 1;$$

One knows that i is an index used for summation, and n represents the number of outcomes of the random variable X, meaning such belongs to the set of natural numbers:

$$i, n \in \mathbb{N} = 1, 2, 3, \cdots$$

and p_i represents the probabilities, which belong to the interval:

$$p_i \in [0, 1]$$

then the amount of uncertainty in X is measured by its *entropy*, defined as:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

Shannon sought a function H

Chapter References

"*All scientific work is based on reference to prior knowledge; innovation is merely an elegant rearrangement of existing ideas.*"
– Albert Einstein (attributed)

[Sh48] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[CC05] C. G. Chakrabarty and I. Chakrabarty, "Shannon Entropy: Axiomatic Characterization and Application," *arXiv preprint*, 2005. DOI: https://doi.org/10.48550/arXiv.quant-ph/0511171.

[Wi18] E. Witten, "A Mini-Introduction to Information Theory," *arXiv preprint*, 2018. Published in *Rivista Nuovo Cimento* (2020). DOI: https://doi.org/10.48550/arXiv.1805.11965, Related DOI: https://doi.org/10.1007/s40766-020-00004-5.

[Ch16] R. X. F. Chen, "A Brief Introduction to Shannon's Information Theory," *arXiv preprint*, 2016. DOI: https://doi.org/10.48550/arXiv.1612.09316.

[TY01] Takuya Yamano, "A Possible Extension of Shannon's Information Theory," *Entropy*, vol. 3, no. 4, pp. 280–292, 2001. DOI: https://doi.org/10.3390/e3040280

[Shl14] Jonathon Shlens, "A Light Discussion and Derivation of Entropy," arXiv preprint arXiv:1404.1998, 2014. Available: https://arxiv.org/abs/1404.1998

[Csi08] I. Csiszár, "Axiomatic Characterizations of Information Measures," *Entropy*, vol. 10, no. 3, pp. 261–273, 2008. DOI: https://doi.org/10.3390/e10030261.

[NVS92] Nambiar, K. K., Varma, P. K., Saroch, V. (1992). "An Axiomatic Definition of Shannon's Entropy," *Applied Mathematics Letters*, 5(4), 45–46. DOI: https://doi.org/10.1016/0893-9659(92)90084-M.

[Sha08] Cosma Shalizi, *Information Theory: Axiomatic Foundations, Connections to Statistics*, Course: 36-350 Data Mining, Carnegie Mellon University, 12 September 2008. Available at: https://www.stat.cmu.edu/~cshalizi/350/2008/lectures/06a/lecture-06a.pdf.

[Chi19] T. Chinburg, *Math 280 - Spring 2019: Entropy and the Mathematics of Evolution*, University of Pennsylvania, 2019. Available at: https://www2.math.upenn.edu/~ted/280S19/math280.html.

[PW16] Y. Polyanski and Y. Wu, *MIT 6.441: Lecture Notes on Information Theory*. MIT OpenCourseWare, 2016. Available at https://ocw.mit.edu/courses/6-441-information-theory-spring-2016/pages/lecture-notes/.

[Khi57] A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications, 1957. Translated by R. A. Silverman. ISBN: 978-0486604343

[HanKob01] T. S. Han and K. Kobayashi, *Mathematics of Information and Coding*, American Mathematical Society, 2001. ISBN: 978-0821805343

[St18] J. V. Stone, "Information Theory: A Tutorial Introduction," *arXiv preprint*, 2018. DOI: https://doi.org/10.48550/arXiv.1802.05968.

[As65] R. B. Ash, *Information Theory*. Dover Publications, 1965. Originally published by Interscience Publishers, 1965. Dover edition, 1990. ISBN: 9780486682105.

[St22] J. V. Stone, *Information Theory: A Tutorial Introduction (Second Edition)*. Sebtel Press, 2022. ISBN: 9781739672706, Ebook ISBN: 9781739672737.

[CT06] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (2nd ed.). Wiley-Interscience, 2006. ISBN-13: 978-0-471-24195-9, ISBN-10: 0-471-24195-4.

[Ma03] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. ISBN: 978-0-521-64298-9.

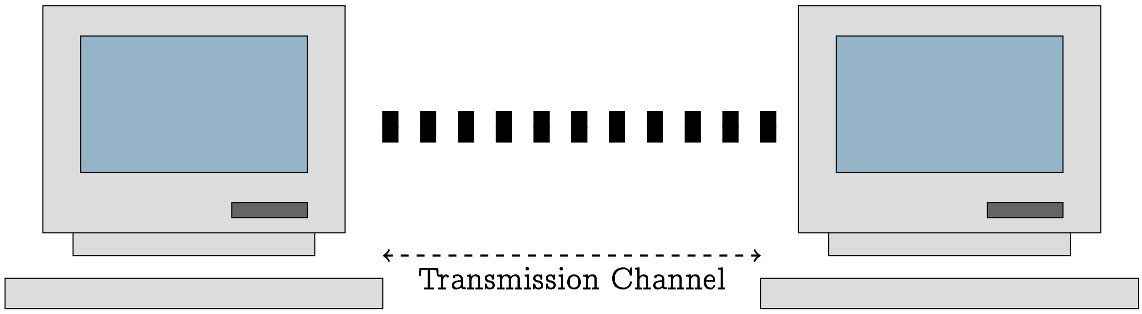
[Gr23] R. M. Gray, *Entropy and Information Theory* (Corrected ed.). 2023. Originally published in 1990 by Springer. DOI: `https://doi.org/10.1007/978-1-4419-7970-4`. Available at: `https://ee.stanford.edu/~gray/it1_corrected_6-23.pdf`.

[EGK12] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012. ISBN: 9781107008731.

XXX to be named later: A General Scenario of Information Transfer

One considers a very general scenario of a communication system. In broad sense, *communication* refers to the transmission of information from a source node to a destination node through a medium. A system like such is designed, so that it ensures the transmitted message is received *accurately* and *efficiently*. At its core, an arbitrary communication system includes some key components such as *information source, transmitter, transmission channel, receiver, and* the *destination*.

As an illustration;



Computer A (Sender / Receiver) Computer B (Receiver / Sender)

One takes two computers; i.e. Computer A and Computer B, the first serving as the information source, and the latter being the destination, both of which are connected by a transmission channel.

[*Input*] → [*Transmitter*] → [*Transmission Channel*] → [*Receiver*] → [*Output*]