

# An overview of the development and applications of information entropy

**Yiqi Huang**

WLSA, Shanghai, 200082, China

yiqihuang484@gmail.com

**Abstract.** With information entropy gradually taking the lead in modern information theory development, it begins to hold greater influence over multiple research areas as well as technology innovation. This paper aims to clarify people's confusion with the development of entropy theory and provide a brief overview of the origin of entropy theory, including the original Shannon's proposal, variants such as relative entropy and conditional entropy, and entropy concepts proposed by other scientists, such as Rényi Tsallis entropy. The paper also includes the current application of entropy, studies hotspots, and predicts future entropy development trends. This research paper is able to add more coherence and consistency to information entropy's development, helping more people to better understand the concept of entropy and its derivation. At the same time, with hotspots of entropy fields of study, this paper hopes to attract more people to devote themselves to studying entropy-related fields, and boost technological development.

**Keywords:** Information entropy, Shannon entropy, Information theory.

## 1. Introduction

Claude E. Shannon first proposed the idea of information entropy in 1948, and it has since grown to be fundamental to the study of thermodynamics, information theory, data analysis, and communication. By effectively quantifying uncertainty and information using mathematical terminology, Shannon's information entropy helped to limit the loss of signal during transmission and storage in the presence of noise.

The goal of this paper is to provide a thorough overview of current information entropy theories and applications. a synopsis of its theoretical foundation, real-world implementation, and continuing research. First, the review covers the foundational ideas of information entropy, along with some basic mathematical formulas and their interpretation. This paper then delves into additional forms of information entropy, which are related ideas that expand on information entropy, including relative entropy, cross-entropy, Rényi entropy, and Tsallis entropy, which are related concepts that extend information entropy to a wider array of contexts and fields. Following that, this paper lists the applications of information entropy in different areas, such as communication systems, machine learning, biological science, and the financial market. In addition, this review discovers current hotspots and development trends of the information entropy study, and explores the potential future direction of the information entropy.

## 2. Fundamental Theory of Entropy

### 2.1. Shannon Entropy

In Claude E. Shannon's paper *A Mathematical Theory of Communication* published in 1948, he provided a mathematical measure of the uncertainty or randomness of a random variable or an information source. Mathematically, Shannon entropy  $H(X)$  for a discrete random variable  $X$  with a probability distribution  $P = \{p_1, p_2, \dots, p_n\}$  is defined as follows [1]:

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

Where  $p_i$  is the probability of occurrence of the  $i$ th possible value of  $X$ , and the logarithm is typically taken in base 2, with entropy measured in bits.

The expected information from a stochastic data source is quantified by the Shannon entropy. For example, when all possible outcomes of a random variable have the same probability, the entropy is maximized, indicating the highest level of uncertainty. On the other hand, when results are certain, there is no uncertainty since the entropy is reduced to zero.

### 2.2. Information and Entropy

Entropy is a term used to describe how unpredictable or random a system is. An increased entropy indicates a more chaotic system, which translates into increased uncertainty and a larger requirement for information to characterize the system's state. Entropy can also be defined as the expected value of a variable's self-information [2]. Entropy plays a fundamental role in information transfer, enabling the least transmission loss through decoding, encoding, and compression.

### 2.3. Joint Entropy and Conditional Entropy

In Shannon's information entropy, he also describes the entropy of a joint event. Suppose there are two events,  $x$  and  $y$ . Then, the joint entropy of the events  $x$  and  $y$   $H(X, Y)$  can be described as following [1]:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \quad (2)$$

Where  $p(x, y)$  is the joint probability of  $X = x$  and  $Y = y$

For the conditional entropy of  $Y$  and  $X$ ,  $H(Y|X)$  is defined as following

$$H(X|Y) = -\sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(y)} \quad (3)$$

Where  $X$  and  $Y$  denote the support sets of  $X$  and  $Y$

The conditional entropy is used to determine the outcome of a random variable  $Y$  given that the other variable  $X$  is known. This measure can be used to understand the dependency between variables.

### 2.4. Relative Entropy (Kullback-Leibler Divergence)

In Kullback's paper *On Information and Sufficiency* published in 1951, he proposed a concept that can be used to measure the difference between two probability distributions  $P$  and  $Q$  [3]:

$$D_{KL}(P||Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} \quad (4)$$

His approach offers a helpful means of figuring out how one probability distribution diverges from another, predicted probability distribution, despite the fact that the KL divergence is not symmetric and is not a genuine metric. This approach is frequently utilized in domains including evaluation, model selection, and machine learning.

### 2.5. Cross Entropy

Cross entropy  $H(P, Q)$  measures the difference between the true probability distribution  $P$  and an estimated probability distribution  $Q$  [4]:

$$H(P, Q) = -\sum_{i=1}^n p_i \log q_i \quad (5)$$

Cross entropy includes both the entropy of the true distribution and the KL divergence, so it can also be written as follows:

$$H(P, Q) = H(P) + D_{KL}(P||Q) \quad (6)$$

Cross entropy is often used in machine learning as a way of a loss function in classification problems to measure the performance of a predictive model.

## 3. Extension and Variants of Entropy

### 3.1. Rényi Entropy

In 1961, Rényi, Alfréd published “On measures of information and entropy”, in his paper, he introduced a quantity that generalized concepts including Hartley entropy, Shannon entropy, collision entropy, and min-entropy. Rényi entropy includes a parameter  $\alpha$  that allows people to adjust the sensitivity to different probability values [5]. The entropy of order  $\alpha$  is defined as follows:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left( \sum_{i=1}^n p_i^\alpha \right) \quad (7)$$

Where  $\alpha \geq 0$  and not equal to 1. If  $\alpha$  approaches  $\infty$ , the entropy would converge to Shannon entropy. The value of  $\alpha$  can be adjusted with the higher the value, the more weight to events with higher probability. Rényi entropy is often involved in areas, such as ecology, where people use Rényi entropy to measure biodiversity.

### 3.2. Tsallis Entropy

Tsallis entropy was introduced in 1988 by Constantino Tsallis as a generalization of the standard Boltzmann–Gibbs entropy. The Tsallis entropy is defined as below [6]:

$$S_q(X) = \frac{1}{q-1} \left( 1 - \sum_{i=1}^n p_i^q \right) \quad (8)$$

Where  $q$  is a real number parameter called entropic-index. The Tsallis entropy would reduce to Shannon entropy as  $q$  approaches 1. This entropy is usually used in the study of non-extensive systems, where standard properties of entropy do not apply.

### 3.3. Other Entropy Measures

Other than the extensions of entropy introduced above, there are other various entropy measures for different needs and applications. Here are some examples:

- (1) Min-entropy: Focusing on most likely events, and defined as  $H_\infty(X) = -\log \max_i p_i$  [7]
- (2) Permutation Entropy: Analyzes time series data by considering the ordinal patterns of the time series values [8].
- (3) Approximate Entropy: Measures the regularity and unpredictability of fluctuations in time series data, useful in analyzing physiological signals [9].

## 4. Applications of entropy

### 4.1. Communication systems

Shannon’s entropy information entropy lays a solid foundation for communication systems and derives multiple methods in data compression and error correction. To specify, data compression means that the information in the message stays unchanged while being represented in fewer characters. Techniques,

such as Huffman coding and arithmetic coding, are able to exploit the principles of information and compress data efficiently [10,11].

At the same time, when processing long-distance communication, data is likely to be affected by outside noise, causing errors. In this case, error correction techniques, such as Reed-Solomon codes and convolutional codes, can add additional bits to help correct the errors [12,13].

#### *4.2. Machine learning and Data Mining*

Information entropy is deeply involved in decision-making algorithms, such as ID3, C4.5, and CART. In those algorithms, their goal is to convey the maximum information gain, which is marked as minimum information entropy, so that more informative and accurate decision trees can be given.

Another feature of machine learning that applies the concept of entropy is the feature selection method, which helps to create more predictive models. By selecting the feature that provides the most information gain, these methods can help accelerate the process of machine learning and improve machine learning models' interpretability [14].

#### *4.3. Biological Sciences*

In the biological sciences, entropy is used to analyze genetic sequences, protein structures, and ecological systems. In genomics, entropy-based measures help identify conserved regions in DNA sequences, indicating functional or evolutionary significance. Entropy is also used in detecting motifs and patterns within a sequence, which benefits gene regulation and expression.

Protein folding also involves entropy changes, where different entropy values are related to different states of the protein conformations and can be helpful in identifying the stable state of the protein as well as its functional states [15].

In ecology, entropy is used to measure the evenness and diversity of the ecological system. Shannon's diversity index is an example where entropy is used to quantify the diversity and abundance of species within an ecosystem [16].

#### *4.4. Finance*

In the financial market, entropy is often used to analyze the market's dynamics and detect anomalies. For example, entropy helps to quantify the unpredictability of market behavior, making investors better understand the market's current situation and risk of investment. Techniques such as approximate entropy and sample entropy are often involved in this case to help companies proceed with risk evaluation and management, and create the most suitable investing strategies based on the results.

#### *4.5. Other fields*

Entropy has applications in various other fields, including: (1) Cryptography: Entropy measures the unpredictability of cryptographic keys, ensuring secure encryption schemes. (2) Linguistics: Entropy is used to analyze the complexity and redundancy of natural languages. (3) Neuroscience: Entropy helps study brain activity patterns, providing insights into neural complexity and information processing. (4) Thermodynamics: Entropy helps to understand thermodynamic systems, energy transfer, and phase changes.

### **5. Current hotspots of entropy study and future research directions**

#### *5.1. Quantum Information Theory*

One of the current hotspots of entropy study is quantum information theory. Currently, in quantum mechanics, entropy-related concepts, for example, von Neumann entropy are involved in research in quantum entanglement, quantum coherence, quantum computing, and quantum communication. To specify, the concept of entropy is crucial in quantum state discrimination, quantum cryptography, and error correction in quantum computers, as mentioned above [17].

### 5.2. Machine learning

The integration of information entropy-based methods and machine learning is a vibrant study subject. Entropy is becoming more and more frequently involved in the process of designing new algorithms for unsupervised learning, such as clustering and anomaly detection. In the current state, since AI is a popular subject, more scholars are joining in the fields to create better models and learning algorithms to improve AI's stability and understanding ability [18].

### 5.3. Network Science

In network science, entropy is used to analyze the complexity and information flow in networks, including social networks, biological networks, and communication networks. Current research in network science is mainly about using entropy to measure community structures, and robustness, and identify crucial nodes for the network's stability. When testifying the resilience and effectiveness of a network system, entropy-based metrics are often involved, particularly in the context of cascading failures and epidemic spreading.

### 5.4. Big Data

In big data, entropy is being used to process data on large scales. For example, researchers would develop entropy-based methods to finish data compression, feature extraction, and detection of errors. These methods are crucial in ensuring the volume, variety, and speed of processing huge amounts of data. Additionally, entropy can be used to improve the security and privacy of data, maintaining a well-protected internet transmitting environment.

### 5.5. Future Directions

The future of entropy research holds several promising directions: (1) Entropy in Artificial Intelligence (AI): Exploring the role of entropy in AI, particularly in enhancing decision-making processes and improving the interpretability of AI models. (2) Advanced Quantum Entropy Measures: Developing new quantum entropy measures and understanding their implications for quantum information processing. (3) Real-Time Entropy Analysis: Creating methods for real-time entropy analysis in dynamic systems, enabling rapid adaptation and decision-making. (4) Entropy and Sustainability: Applying entropy concepts to address environmental challenges, such as resource management and sustainability.

## 6. Conclusion

Even since Claude Shannon proposed the concept of information entropy in 1948, the variety of entropy has kept increasing. Scientists have come up with different variations of entropy that suit different situations and are used in different applications. This review has highlighted the fundamental aspects of entropy, including the basic mathematical definition of Shannon entropy, and its interpretation. Then, the review introduces other basic variants of entropy, such as cross entropy, relative entropy, joint entropy, and conditional entropy. Following that, the Rényi entropy, and Tsallis entropy are included, which further border on the application of entropy.

From ecology to machine learning, this review has introduced multiple applications that use each type of entropy. In communication systems, entropy underpins data compression and error correction techniques that are essential in data transmission. In machine learning, entropy-based methods help improve models and algorithms' stability and interpretability, aiding people to create better models that can be used in network systems, processing big data, and artificial intelligence. In biology, entropy is used to help interpret genetic sequence, helping to understand protein folding's function and stable states.

With those applications, current entropy mainly has several popular study fields. In quantum information theory, entropy is used to understand further quantum-level interactions. In machine learning, entropy helps to improve models. Other than that, there are still many areas, including big data and artificial intelligence, that are still discovering the use of entropy. On top of that, the future direction

of entropy studies is pointing toward broader applications such as AI, sustainability, and real-time dynamic system analysis.

To sum up, information entropy is the basis of modern science and technology and provides a special perspective for observing and understanding the complexity of systems. With the development of technology, the concept of entropy will certainly become a useful tool in understanding our surrounding world, and will have more applications in people's daily lives.

## References

- [1] Shannon, Claude E. (1948) A Mathematical Theory of Communication. Bell System Technical Journal. July, 27 (3): 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
- [2] Pathria, R. K.; Beale, Paul. (2011) Statistical Mechanics (Third ed.) Academic Press, pp. 51.
- [3] Kullback, S.; Leibler, R.A. (1951) On information and sufficiency. Annals of Mathematical Statistics. 22 (1): 79–86. doi:10.1214/aoms/1177729694.
- [4] Thomas M. Cover; Joy A. Thomas (2006) Elements of Information Theory. Hoboken, New Jersey: Wiley, July 18.
- [5] Rényi, Alfréd. (1961) On measures of information and entropy. Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960, pp. 547–561.
- [6] Tsallis, C. (1988) Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics. 52 (1–2): 479–487. doi:10.1007/BF01016429.
- [7] Vazirani, Umesh; Vidick, Thomas (2014). Fully Device-Independent Quantum Key Distribution. Physical Review Letters. September 29, 113 (14): 140501. doi:10.1103/physrevlett.113.140501.
- [8] Aczél, J.; Forte, B.; Ng, C. T. (1974). Why the Shannon and Hartley entropies are 'natural'. Advances in Applied Probability. 6 (1): 131–146. doi:10.2307/1426210.
- [9] Pincus, S. M.; Gladstone, I. M.; Ehrenkranz, R. A. (1991) A regularity statistic for medical data analysis. Journal of Clinical Monitoring and Computing. 7 (4): 335–345. doi:10.1007/BF01619355.
- [10] Huffman, D. (1952) A Method for the Construction of Minimum-Redundancy Codes. Proceedings of the IRE. 40 (9): 1098–1101. doi:10.1109/JRPROC.1952.273898.
- [11] MacKay, David J.C. (2003). Chapter 6: Stream Codes. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, September. Archived from the original (PDF/PostScript/DjVu/LaTeX) on 22 December 2007.
- [12] Gorenstein, D.; Zierler, N. (1961) A class of cyclic linear error-correcting codes in  $p^m$  symbols, June. J. SIAM. 9 (2): 207–214. doi:10.1137/0109020.
- [13] Benedetto, Sergio, and Guido Montorsi. (1995) Role of recursive convolutional codes in turbo codes. Electronics Letters, 31.11: 858-859.
- [14] Sarangi, Susanta; Sahidullah, Md; Saha, Goutam (2020) Optimization of data-driven filterbank for automatic speaker verification. Digital Signal Processing, September, 104: 102795. arXiv:2007.10729.
- [15] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walters P. (2002) The Shape and Structure of Proteins. Molecular Biology of the Cell; Fourth Edition. New York and London: Garland Science.
- [16] Spellerberg, Ian F., and Peter J. Fedor. (2003) A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the “Shannon–Wiener” Index. Global Ecology and Biogeography 12.3, 177-179.
- [17] Helstrom, Carl W. (1976) Quantum detection and estimation theory. New York: Academic Press. ISBN 978-0-12-340050-5. OCLC 316552953.
- [18] Rubinstein, Reuven Y.; Kroese, Dirk P. (2013). The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning. Springer Science & Business Media, March 9.