Acta Scientific COMPUTER SCIENCES

Volume 3 Issue 10 October 2021

Review Article

A Profile of the Generalized Information Measures in Information Theory

D S Hooda[1]* and M S Barak[2]

[1]*Honorary Professor in Mathematics, GJ University of Science and Technology, India*

[2]*Department of Mathematics, IG University, Meerpur, Rewari, India*

***Corresponding Author:** D S Hooda, Honorary Professor in Mathematics, GJ University of Science and Technology, India.

Received: August 11, 2021

Published: September 29, 2021

Abstract

Information theory deals mainly with Entropy or Information Measure, Communication and Cryptography. A review of information measures with their historical development is an important input in Information theory. The concept of Shannon's entropy and its properties are very informative. The application of entropy in coding is described with an example. Various generalizations of entropy by various authors are enumerated. The 'useful' information measure is defined with its generalization. The measures of Directed divergence and J-divergence are discussed. The 'useful' relative information and j-divergence measures are also described with conclusion and discussion in the end.

Keywords and Phrases:: Entropy; Information Measure; 'Useful' Information Measure; Directed Divergence and J-divergence.

Historical Development of Information Theory

Information theory was usually considered as a branch of mathematics, however, that was made mathematically rigorous only in 1940's. The term 'information theory" does not possess a unique definition, broadly speaking information theory deals with the study of problems concerning information about any system or process. This includes information processing, information storage, information retrieval and decision making. In a narrow sense, information theory studies all theoretical problems connected with the transmission of information over communication channels. This includes the study of information measures and various practical and economical methods of coding information for transmission. It also includes security of information and error correcting codes.

The first study in this direction was undertaken by Nyquist [33] and Hartley [13] who recognized the logarithmic nature of the measure of information. Nyquist's paper contains a theoretical section quantifying "intelligence" and the " line speed" at which it can be transmitted by a communication system, giving the relation W=K log m, where W is the speed of transmission, m is the number of different voltage levels to choose from at each time step, and K is a constant. Ralph Hartely's paper uses the word information as measurable quantity, reflecting the receiver's ability to distinguish that one sequence of symbols from any other, thus quantifying information as $H=\log S^n= n \log S$, where S is the number of possible symbols, and n is the numbers of symbols in a transmission.

Information theory as an independent discipline came into existence about 72 years back since the publication of the first seminal paper on the mathematical theory of communication by Shannon [40], this subject invited the attention of transmitted signal and in the type of decision made at the receiver. In the Shannon model messages are first encoded and then transmitted, whereas in the

Wiener model the signal is communicated directly through the channel without being encoded.

In the sequential years, Shannon studied the applications of information theory in many fields, such as physics, urban and regional planning, marketing, spectral analysis, computerized tomography, economics, etc. This was also interpreted as measure of equality and was used as in linguistics, Psychology and other subjects of social and management sciences. In a large of applications the proportions are treated as probabilities with their sum as unity. Consequently, information theory developed the interest to the scientist of physical, social and biological sciences as well as the researchers in humanities.

Jaynes's [23] work was another major step in developing the scope of information theory apart from communication theory where he commenced the principle of maximum entropy and applied it to statistical mechanics. He showed by using maximum entropy that all the results of statistical mechanics could be developed in a simple manner.

Kullback [27] dealt with the information theory from statistical point of view and that created the interest of statisticians in information theory, but this did not happen in a big way. Statisticians continued taking minimal interest in information theoretic approach to their areas of research. Theil [48] amalgamated the entropy concept with the measure of economic inequalities. He used the entropy function to measure income inequality and industrial concentration.

In the present paper the concept of Shannon's entropy and its properties are explained in section 3 and the application of entropy in coding is described in section 3.Various generalizations of entropy are given in section 4. In section 5, the 'useful' information measure is defined, and its generalization is studied in section 6.In section 7, the measures of Directed divergence and J-divergence are discussed .The 'useful' relative information and j-divergence measures are described in section 8 with conclusion and discussion in section 9. An exhaustive list of references is given at the end of the paper.

Concept of shannon entropy and properties

The word entropy is a much overused word in today's scientific world. Although it originated in the literature on thermodynamics, yet it has been used in many disciplines because of its association with the concepts of information as envisaged by Claude Shannon [40] in the mathematical theory of communication. A great deal of insight is obtained by considering it equivalent to uncertainty, which plays a very significant role in our different perceptions about the external world. Any discipline that can assist us in understanding it, measuring it, maximizing or minimizing it, and ultimately controlling it to the extent possible, should be considered an important contribution to our scientific understanding of complex phenomena.

As our perceptions of the world become intensively complex, the number of phenomena about which we are uncertain and the uncertainty in each phenomenon can also increase. To decrease this uncertainty, we collect an ever-increasing amount of information. Uncertainty is not a single-meaning concept. It can appear in several guises. For instance, it can arise in a probabilistic phenomenon such as the tossing of a coin or a die. On the other hand, it can also appear in a deterministic phenomenon where we know the outcome is not a chance event; we are fuzzy about the probability of the specific outcome. The former is called probabilistic uncertainty, while the latter is known as fuzzy uncertainty. Here we study the probabilistic uncertainty.

There may be uncertainty about a candidate winning an election, which brand of a product will have the maximum share of the market?, what grade a student will receive? There may be n possible outcomes in each one of the situations just cited and their probabilities may be $p_1, p_2, p_3, \ldots\ldots\ldots\ldots, p_n$, where $p_i \geq 0$ for each *i* and $\sum_{i=1}^{n} p_i = 1$. Here, we are uncertain as to which outcome will be realized though we know the values $p_1, p_2, p_3, \ldots\ldots\ldots\ldots, p_n$.

Different probability distributions have different uncertainties associated with them e.g. a probability distribution (0.5, 0.5) has much more uncertainty than (0.0001, 0.9999). Can we find a measure of how much uncertain we are about the outcome of a probability distribution? Let $H(p_1, p_2, p_3, \ldots\ldots\ldots\ldots, p_n)$ be a measure, then it is reasonable to require of it the following properties:

- H should be a continuous function of p_i's.
- If all p_i's are equal i.e. $p_i = 1/n$ for each *i*, then H should be a monotone increasing function of n.

- 3. $H(p_1, p_2, p_3, \ldots\ldots\ldots, p_n) = H(p_1 + p_2, p_3, \ldots\ldots\ldots, p_n)$
  $(p_1 + p_2) H(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$

Shannon (1948) proved that the only function H satisfying these axioms is of the form:

$$H(p) = -k \sum_{i=1}^{n} p_i \log p_i \text{ ,----(2.1)}$$

Where k is a positive constant and is determined by taking $H(1/2, 1/2) = 1$.

Thus k=1 and (1.1) is a measure of information or uncertainty removed. However, it was named entropy on the advice of his friend Johan Van Neumann.

Properties of shannon's entropy

The Shannon's entropy satisfies the following properties:

- $H_n(p)$ is permutation ally symmetric i.e. it does not change if $p_1, p_2, p_3, \ldots\ldots\ldots, p_n$ are re-ordered among themselves.
- $H_n(p)$ is a continuous function of p_i when $0 < p_i < 1$. At $p_i = 0$, $-p_i \log p_i$ is not defined but $\lim_{p\to 0} -p \log p = 0$. In case we define $0.\log 0 = 0$, then $H(p)$ is also continuous at $p_i = 0$.
- The entropy does not change by the inclusion of an impossible event i.e. $H_{n+1}(p_1, p_2, \ldots\ldots p_n, 0) = H_n(p_1, p_2, \ldots\ldots p_n)$ i.e.
- If $0 < p_i < 1$, then $\log p_i < 0$ and $-\log p_i > 0$; further if $p_i = 0$ or 1, then $p_i \log p_i = 0$ and that follows $-p_i \log p_i \geq 0$ or $-p_i \sum \log p_i \geq 0$.
- $H(p) = 0$ for all degenerate distributions.
- Let $\varphi(p) = -p \log p$, then it can be verified that $\varphi(p)$ is a concave function. Since the sum of concave functions is also a concave function, therefore $H_n(P)$ is a concave function of $p_1, p_2, p_3, \ldots\ldots\ldots, p_n$.
- $H_n(p)$ is maximum for $P = \left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots\ldots\ldots, \frac{1}{n}\right)$.
- Let $P = (p_1, p_2, p_3, \ldots\ldots\ldots p_n)$ and $Q = (q_1, q_2, q_3, \ldots\ldots\ldots q_n)$ be two independent probability distributions of two random variables X and Y respectively, so that $p(X = x_i) = p_i$, $p(Y = y_i) = q_j$ and $p(X = x_i, Y = y_j) = p_i q_j$. The joint distribution of X and Y is denoted by $P * Q$. Then $H_{mn}(P * Q) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_j \log p_i q_j = H_n(P) + H_m(Q)$.

This is called the additive property of the measure of entropy.

If P and Q are not independent, then

$P(X = x_i) = p_i, P(Y \neq y_j / X = x_i) = q_{ij}$ so that $P(X = x_i, Y = y_j) = p_i q_{ij}$ and

$$H_{mn}(P * Q) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_{ij} \log p_i q_{ij}$$

$$= -\sum_{i=1}^{n} (p_i \log p_i) \sum_{j=1}^{m} q_{ij} - \sum_{i=1}^{n} p_i \sum_{j=1}^{m} q_{ij} \log q_{ij}$$

$$= H_n(P) + \sum_{i=1}^{n} p_i H(Q_i), \text{------ (2.2)}$$

Where $H(Q_i) = -\sum_{j=1}^{m} q_{ij} \log q_{ij}$ is the entropy of conditional probability distribution of Y random variable when X has resulted in the i^{th} outcome. (1.2) can also be written as

$$H_{mn}(P * Q) = H_m(Q) + \sum_{j=1}^{m} q_j H(P_j) \text{ ---------(2.3)}$$

For any converse function $f(x)$, we have by definition

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i) \text{ ----(2.4)}$$

Let $f(x) = x \log x$ be a converse function and $x_i = q_{ij}$, then

$$f\left(\sum_{i=1}^{n} p_i q_{ij}\right) = \left(\sum_{i=1}^{n} p_i q_{ij}\right) \log\left(\sum_{i=1}^{n} p_i \log q_{ij}\right) \leq \sum_{i=1}^{n} p_i q_{ij} \log q_{ij} \text{ -------(2.5)}$$

Now

$$\sum_{i=1}^{n} p_i q_{ij} = \sum_{i=1}^{n} P(X = x_i) P(Y = y_j / X = x_i) = \sum_{i=1}^{n} P(X = x_i, Y = y_j) = q_j$$

Putting this in (1.2), we have

$$q_j \log q_j \leq \sum_{i=1}^{n} p_i q_{ij} \log q_{ij} \text{ ----- (2.6)}$$

Taking Summation over j from 1 to m, we get

Combining (1.2) and (1.7), we have

$H_{mn}(P * Q) \leq H_n(P) + H_m(Q)$, which is called the sub-additive property.

$$H_n(p_1, p_2, \ldots\ldots\ldots, p_n) = H_{n-1}(p_1 + p_2, p_3, \ldots\ldots\ldots, p_n) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Hence, $H_{n-1}(p_1 + p_2, p_3, \ldots\ldots\ldots, p_n) \leq H_n(p_1, p_2, p_3, \ldots\ldots\ldots, p_n)$.

Thus if two outcomes are combined, the entropy is reduced. This property is again desirable since when two outcomes are combined the uncertainty should not increase.

$$H_n(P) = H_{n-1}(p_1 + p_2, p_3, \ldots\ldots\ldots, p_n) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Application of entropy in coding

When we send messages in English language through certain communication channels, we have first to code these in terms of code alphabets. We may use a binary code with two alphabets 0 and 1 or a code consisting of D alphabets.

Suppose we have code words which have lengths $l_1, l_2, \ldots\ldots, l_n$ and their probabilities are $p_1, p_2, \ldots\ldots\ldots, p_n$ then the expected length of the encoded message is $L = \sum_{i=1}^{n} p_i l_i$. thus suppose we are using the binary code with alphabets 0 and 1, then the code words available are.

Message symbols	Probability	Code words	Codeword length
x_1	p_1		
x_2		0	1
x_3	p_2	1	1
x_4	p_3	00	2
x_5	p_4	01	2
x_6	p_5	10	2
....	p_6	11	2
...			

Table 1: Probability and codeword lengths of a Message.

The expected codeword lengths of the best 1-1 code is given by

$L_{1:1} = (p_1 + p_2)\,1 + (p_3 + p_4 + p_5 + p_6)\,2 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

However in general we may have another restriction on $l_1, l_2, \ldots\ldots., l_n$ since we have not only to code messages, but we have also to decode them and we want to design a code in such a way as to ensure unique decipherability. The necessary and sufficient condition for this was given by Kraft in the master's thesis of electrical Engineering and the condition known as Kraft's inequality is $\sum_{i=1}^{n} D^{-l_i} \le 1$.

Shannon minimized $\sum_{i=1}^{n} p_i l_i$ subject to $\sum_{i=1}^{n} D^{-l_i} \le 1$ and proved a theorem

$$H_D(P) \le L < H_D(P) + 1.$$

For the best uniquely decipherable code $l_i = -\log_D p_i$ for each i is the minimum codeword length and $l_i \ge -\log_D p_i$.

Various generalizations of entropy

In this section we give the definitions of parametric and non-Parametric entropies in the following table 2.

Sr. No.		Definition
Non-Parametric		
1.	Shannon [40]	$H(P) = -\sum_{i=1}^{n} p_i \log p_i$.
2.	Pal and Pal [40]	$E(P) = \frac{1}{n(\sqrt{e}-1)} \sum_{i=1}^{n} p(x_i)\left(e^{(1-p(x_i))} - 1\right)$.
Parametric		
3.	Renyi [39]	$H_\alpha(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha\right), \alpha \ne 1, \alpha > 0.$
4.	Havrada and Charvat [14]	$H^\alpha(P) = \frac{1}{2^{1-\alpha}-1}\left(\sum_{i=1}^{n} p_i^\alpha\right) - 1,\ \alpha \ne 1, \alpha > 0$.

5.	Kapur [24]	$H_{\alpha,\beta}(P)=\frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n}p_i^{\alpha+\beta-1}\Big/\sum_{i=1}^{n}p_i^{\beta}\right),\alpha\neq 1,\alpha>0,\beta>0$, $\alpha+\beta\text{-}1>0$.
6.	Kapur [25]	$H_{\alpha,\beta}(P)=\frac{1}{\beta-\alpha}\log\frac{\sum_{i=1}^{n}p_i^{\alpha}}{\sum_{i=1}^{n}p_i^{\beta}},\alpha\neq\beta$.
7.	Behara and Chawla [4]	$H_{\tau}(P)=\frac{1-\left(\sum_{i=1}^{n}p_i^{1/\tau}\right)^{\tau}}{1-2^{\tau-1}},\tau\neq 1,\tau>0.$
8.	Sharma and Mittal [24]	$H^{\beta}(P)=(2^{(\beta-1)\sum_{i=1}^{n}p_i\log p_i})(2^{1-\beta}-1)^{-1},\beta>0,\ \beta\neq 1$; $H_{\alpha}^{\beta}(P)=\left[(\sum_{i=1}^{n}p_i^{\alpha})^{\frac{\beta-1}{\alpha-1}}-1\right](2^{1-\beta}-1)^{-1},\alpha\neq\beta,\ \alpha>0,\ \beta>0,\ \alpha\neq 1.$
9.	Arimoto [3]	${}_{t}H(P)=\frac{1}{2^{t-1}-1}\left[\left(\sum_{i=1}^{n}p_i^{\frac{1}{t}}\right)^{t}-1\right];\ t\neq 1,t>0$.
10.	Boekee and Lubbe [7]	$H_R(P)=\frac{R}{R-1}\left[1-(\sum_{i=1}^{n}p_i^{R})^{\frac{1}{R}}\right],R>0,R\neq 1.$
11.	Kapur [25]	$H_{\alpha}(P)=-\sum_{i=1}^{n}p_i\log p_i+\frac{1}{a}\sum_{i=1}^{n}(1+ap_i)\log(1+ap_i)-\frac{1}{a}(1+a)\log(1+a)$, $\alpha\geq 1$.
12.	Kapur [26]	$H_{\alpha,\beta}(P)=\frac{1}{\alpha+\beta-2}\left[\sum_{i=1}^{n}p_i^{\alpha}+\sum_{i=1}^{n}p_i^{\beta}-2\right],\alpha\neq\beta$

13.	Sharma and Taneja [42]	$H_\alpha^\beta(P)=\frac{1}{\beta-\alpha}\left[\sum_{i=1}^n p_i^{\alpha}-\sum_{i=1}^n p_i^{\beta}\right],\alpha\neq\beta$
14.	Kvalseth [31]	$E_\alpha(P)=\frac{\sum_{i=1}^n p(x_i)\left(e^{\left(1-p^\alpha(x_i)\right)}-1\right)}{\alpha},\alpha>0$
15.	Hooda and Ram [16]	$H_R^\beta(P)=\frac{R}{R+\beta-2}\left[1-\left(\sum_{i=1}^n p_i^{\frac{R}{2-\beta}}\right)^{\frac{2-\beta}{R}}\right];0<\beta\leq 1,R>0\,and$ $R+\beta\neq 2$.
16.	Hooda and Sharma [17]	$H_R^{(\alpha,\beta)}(P)=\frac{R}{R+\beta-2\alpha}\left[1-\left(\sum_{i=1}^n p_i^{\frac{R}{2\alpha-\beta}}\right)^{\frac{2\alpha-\beta}{R}}\right];$ $\alpha\geq 1,\ 0<\beta\leq 1, R(>0)\neq 1,\ R+\beta\neq 2\alpha$.
17.	Kumar and Hooda [29]	$H_R^\alpha(P)=\frac{R}{R-\alpha}\left[1-\left(\sum_{i=1}^n p_i^{\frac{R}{\alpha}}\right)^{\frac{\alpha}{R}}\right];\ 0<\alpha\leq 1, R(>0)\neq 1.$
18.	Kumar., *et al.* [30]	$H_R^m(P)=\frac{R-m+1}{R-m}\left[1-\left(\sum_{i=1}^n p_i^{R-m+1}\right)^{\frac{1}{R-m+1}}\right];R-m+1>0,$ $R\neq m,\ R,m>0(\neq 1)$

Table 2: Definitions of Parametric and Non-Parametric Entropies.

'Useful' information measures

The quantity (1.1), in some sense, measures the amount of information of the probability distribution P. However, this measure does not take into account the effectiveness or importance of the events, while in some practical situations of probabilistic nature subjective considerations also play their role. Taking into account the effectiveness of the outcomes, Belis and Guisau [5] introduced utility distribution $U=(u_1,u_2.,......................u_n)$, where $u_i>0$, is the utility of i^{th} an event having the probability of occurrence p_i.

The occurrence of an event E_i with probability p_i brings to the observer a quantity of information measured by its self-information $H(P_i) = -\log_2 p_i$. Belis and Guisau [5] considered the problem of constructing an analogous function for measuring the amount of self useful information given $H(p_i u_i) = -u_i \log_2 p_i$ based on the following two postulates.

B_1. If all the events of the random experiment having the same utility u>0 say, then the self-useful information provided by the logical product $E_1 \cap E_2$ of two statistically independent events E_1 and E_2 is the sum of the self-useful information provided by E_1 and E_2. In other words,

$$H(p_1 * p_2; u) = H(p_1; u) + H(p_2; u) \quad \text{---- (5.1)}$$

Where $p_1 * p_2$ is the probability of $E_1 \cap E_2$.

B_2. The self-useful information provided by the occurrence of E_1 is proportional to its utility $u_i > 0$, i.e.; for each $\delta > 0$,

$$H(p_i; \delta u_i) = \delta H(p_i; u_i) \quad \text{------(5.2)}$$

The mean value of self-useful information to the probability distribution P was given by Belis and Guisau [5] as $\Phi_0(P)$:

$$H(P;U) = -\sum_{i=1}^{n} u_i p_i \log_2 p_i ; u_i > 0 \sum_{i=1}^{n} p_i = 1, \quad \text{---------(5.3)}$$

Which was called 'useful' information by Longo [32] and weighted entropy by Guiasu and Picard [12].

On defining preference as a mapping of utility distribution over probability distribution, i.e.,

$u_i p_i = v_i, f: U * P \to V$, Such that $f(u_i p_i) = v_i$, (4.3) was modified as

$$H(P;U) = -\sum_{i=1}^{n} v_i \log p_i ; where \quad v_i = u_{i_i} p_i \forall i \cdot \quad \text{------- (5.4)}$$

It is a generalization of Shannon's entropy as it reduces to $\Phi_1(P)$ where $u_i = 1$ each i or utilities are ignored.

Generalized 'useful' information measures

Belis and Guiasu's measure given by (5.3) satisfies the additive property of the type

$$H(P * Q; U * V) = \overline{V} H(P;U) + \overline{U} H(Q;V), \quad (6.1)$$

Where

$$P * Q = (p_1 q_1, p_2 q_2, \ldots\ldots p_n q_m \ldots\ldots; p_n q_1, p_n q_2 \ldots\ldots p_n q_m)$$

$$U * V = (u_1 v_1, u_1 v_2, \ldots\ldots u_1 v_m; \ldots\ldots; u_n v_1, u_n v_2 \ldots\ldots u_n v_m)$$

$\overline{V} = \sum_{j=1}^{m} q_j v_j$ and $\overline{U} = \sum_{i=1}^{n} p_i u_i$

Emptoz [10] considered a non-additive generalization of (6.1). Aggrawal and Picard [2] characterized the following generalized measure:

$$H_\beta(P;U) = \frac{\sum_{i=1}^{n} u_i p_i (p_i^{\beta-1} - 1)}{2^{1-\beta} - 1} ; \beta \neq 1, \beta > 0$$

and called it entropy of type β with preference.

Hooda and Tuteja [19] generalized (5.1) by considering.

$$H(P * Q; U * V) = (\sum_{j=1}^{m} v_j q_j)^\alpha H(P;U) + (\sum_{i=1}^{n} u_i p_i)^\alpha H(Q;V) \quad \text{---- (6.2)}$$

And obtained the following measure of useful information

$$H(P,U) = -2^{\alpha-1} \sum_{i=1}^{n} (u_i p_i)^\alpha \log p_i ; \alpha \neq 0 \quad \text{----- (6.3)}$$

They also considered a non-additive generalization of (6.2) and obtained the measure of useful information

$$H(P;U) = \frac{\sum_{i=1}^{n} u_i^\alpha p_i^\alpha (p_i^{\beta-\alpha} - 1)}{2^{\beta-1} - 2^{\alpha-1}} ; \alpha, \beta > 0 \text{ and } \beta \neq \alpha \quad \text{------- (6.4)}$$

Sharma, Mohan, and Mitter [43] also characterized two measures of useful information for incomplete probability distributions. They also obtained moments of MLE of (5.3) while Hooda and Kumar [21] and studied moments of MLE of (6.3) and (6.4).

On taking powers α *and* β of $\sum_{j=1}^{m} q_j v_j$ *and* $\sum_{i=1}^{n} p_i u_i$, Taneja and Hooda [46] characterized:

$$H(P,U) = (2^{\beta-1} - 2^{\alpha-1})^{-1} \sum_{i=1}^{n} [(u_i p_i)^\alpha - (u_i p_i)^\beta] \quad \alpha, \beta > 0 \quad \alpha \neq \beta \quad \text{-- (6.5)}$$

Taneja (47) considered the generalized additive relation of the following form:

$$H(P*Q;U*V)=H(P;U)G(Q;V)+G(P;U)H(Q;V) \text{ ----(6.6)}$$

And obtained the following three measures of useful information:

$$H(P,U)=-2^{\alpha-1}\sum_{i=1}^{n}[p_i^{\alpha}u_i^{\beta}\log p_i;\alpha>0,\beta\geq 0 \quad (6.7)$$

$$H(P,U)=(2^{1-\alpha}-2^{1-\beta})^{-1}\sum_{i=1}^{n}u_i^{\beta}(p_i^{\alpha}-p_i^{\gamma}),\alpha,\gamma>0,\alpha\neq\gamma,\beta>0 \text{ -----(6.8)}$$

$$H(P,U)=-2^{\alpha-1}\quad /\sin\gamma\sum_{i=1}^{n}[p_i^{\alpha}u_i^{\beta}\sin(\gamma\log p_i):\alpha>0,\gamma\neq 0,\beta\geq 0 \text{ ----- (6.9)}$$

The measure (5.3) was studied and generalized for complete probability distribution by many authors. However, this is not an additive measure and has a widely spread range which is sometimes difficult to manage.

To overcome these deficiencies, Bhaker and Hooda [6], gave the generalized mean value characterization of the following useful information measures for incomplete probability distributions:

$$H(P;U)=-\frac{\sum_{i=1}^{n}u_ip_i\log p_i}{\sum_{i=1}^{n}u_ip_i} \text{ -------- (6.10)}$$

And

$$H_{\alpha}(P;U)=\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n}u_ip_i^{\alpha}}{\sum_{i=1}^{n}u_ip_i};\alpha\neq 1,\alpha>0 \text{ ------ (6.11)}$$

They have also studied their properties and application in noiseless coding theory. If we take u_i = 1, for each i, (6.10) and (6.11) reduce to entropies of order 1 and α respectively.

It is worth mentioning that Hooda and Singh [18] characterized the following generalized useful information measure of order α and type β through information generating function:

$$H_{\alpha}^{\beta}(P;U)=\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n}u_i^{\beta}p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}};\alpha\neq 1,\alpha>0,\beta>0 \text{ ----(6.12)}$$

Aggarwal and Picard [2] characterized another generalized measure as given below:

$$H_{\beta}(P,U)=\frac{\sum_{i=1}^{n}u_ip_i\left(p_i^{\beta-1}-1\right)}{2^{\beta-1}-1},\ \beta\neq 1,\ \beta>0, \text{ -------(6.13)}$$

Which was called the generalized entropy of type β with preference.

Measures of relative information and J-Divergence

The most important measure of directed divergence of probability distribution P and Q given by Kullback and Leibler [28] is

$$D(P;Q)= p_i\log\frac{p_i}{q_i}, \qquad \text{-----------(7.1)}$$

Where P= {p_1,p_2,....p_n} and Q={q_1,q_2,....q_n} are the probability distributions and pi is considered zero whenever qi=0 and $0\log\frac{0}{0}=0.$

The Kulback-Leibler measure (7.1) has been generalized by Aczel and Nath [1] Patni and Jain [35] and Hooda and Tuteja [20]. It may be noted that the measure (7.1) is not symmetric in P and Q so Kulback [27] defined the following measure and called it.

J-divergence:

$$J(P/Q)=\sum_{1=1}^{n}p_i\log\frac{p_i}{q_i}+\sum_{1=1}^{n}q_i\log\frac{q_i}{p_i},\text{for all }P,Q\in\Delta_n, \text{ ------(7.2)}$$

Where $\Delta_n=\left\{P=(p_1,p_2,\ldots,p_n):p_i>0,\sum_{i=n}^{n}p_i=1\right\},$

The measure (6.2) has found many applications in statistics [Ref. to Kulback [27]] and in pattern recognition.

This measure (7.2) can also be written as

$$J(P/Q)=D(P/Q)+D(Q/P). \text{ -------(7.3)}$$

Renyi [39] gave a scalar parametric generalization of Kulback-Leibler's directed- divergence measure as given below:

$$H^{\alpha}(P:Q)=(\alpha-1)^{-1}\left(\sum_{i=1}^{n}p_i^{\alpha}q_1^{1-\alpha}\right),,\alpha\neq 1,\alpha>0. \text{ ------(7.4)}$$

Sharma and Autar [41] gave another generalization of (7.1)

$$H_{\beta}^{1}(P:Q)=(\beta-1)^{-1}\left(\sum_{i=1}^{n} p_i^{\beta} q_1^{1-\beta}-1\right),, \beta \neq 0,1. \text{ ------(7.5)}$$

In particular, we have

$$\lim_{\alpha\to 1} H_{\beta}^{\alpha}(P:Q)=\lim_{\beta\to 1} H_{\beta}^{1}(P:Q)=K(P:Q) \text{ ------(7.6)}$$

And

$$\lim_{\beta\to 1} H_{\beta}^{2}(P:Q)=H(P:Q) \text{ and } \lim_{s\to 0} H_{s}^{2}(P:Q)=H(P:Q). \text{ -------(7.7)}$$

Rathie and Kanappan [37] gave another generalization of (7.1) as given below.

$$I^{\beta}(P:Q)=\left(\sum_{i=n}^{n} p_i^{\beta} q_{\beta}^{1-\beta}-1\right)\left(2^{1-\beta}-1\right) \text{ for } \beta \neq 1, \text{ (7.8)}$$

Which was called the generalized measure of directed-divergence of order β.

Rathie and Sheng [38] and Burbea and Rao [8] studied the following generalization of degree β given by:

$$J_{\beta}(P/Q)=\left(2^{\beta-1}-1\right)\left\{\sum_{i=1}^{n} p_i^{\beta} q_i^{\beta-1}+\sum_{i=1}^{n} q_i^{\beta} p_i^{\beta-1}-2\right\} \text{ ------------(7.9)}$$

$$=I^{\beta}(P/Q)+I^{\beta}(Q/P), \text{ -------(7.10)}$$

Where $I^{\beta}(P/Q)=\left(2^{\beta-1}-1\right)^{-1}\left[\sum_{i=n}^{n} p_i^{\beta} q_i^{1-\beta}-1\right]$ and

$$I^{\beta}(Q/P)=\left(2^{\beta-1}-1\right)^{-1}\left[\sum_{i=n}^{n} q_i^{\beta} p_i^{1-\beta}-1\right],$$

In case the predicted probability distribution is revised to $R=(r_1,r_2,\dots,r_n)\in\Delta_n$, then improvement in relative information from P to Q is

$$I(P/Q/R)=I(P/Q)-I(P/R)$$

$$=\sum_{i=1}^{n} p_i \log\left(\frac{r_i}{q_i}\right) \text{ -------(7.11)}$$

The measure (6.11) was firstly defined by Theil [48] and was called Theil's measure of information improvement and had many applications in economics.

'Useful' relative information and J-divergence

Let $\Delta_n=\left\{P=(p_1,p_2,p_3,\dots\dots\dots,p_n), 0<p_i\leq 1 \text{ and } \sum_{i=1}^{n} p_i=1\right\}$ be a family of probability distributions for $n\geq 2$ defined on random variable X.

Kullback and Leibler [28] defined the directed divergence or discrimination measure. However, it does not take into account important or utility of events of the experiment. Considering qualitative aspect of events Hooda and Tuteja [20] proposed:

$$D(P\mid Q;U\mid V)=\frac{\sum_{i=1}^{n} w_i p_i \log(p_i/q_i)}{\sum_{i=1}^{n} p_i} \text{ -----(8.1)}$$

Where $U=(u_1,u_2,\dots\dots..u_n)$ and $V=(v_1,v_2,\dots\dots\dots,v_n)$ are utility distributions attached to P and Q respectively and $w_i=\frac{u_i}{v_i}$ for each $i=1,2,3,\dots..,n$. The measure (8.1) was called 'useful' relative information measure.

Later on considering utility of an event independent of probability Taneja and Tuteja [45] proposed and characterized the following measure for complete discrete probabilities distributions:

$$D(P\mid Q;U)=\sum_{i=1}^{n} u_i\, p_i \log\frac{p_i}{q_i} \text{ ---------(8.2)}$$

Bhaker and Hooda [6] defined and characterized 'useful' relative information measure given below:

$$D(P\mid Q;U)=\frac{\sum_{i=1}^{n} u_i\, p_i \log\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} u_i p_i} \text{ ------(8.3)}$$

The measure (8.3) satisfies and the additive property:

$$D(P*R:Q*S;U*V)=D(P:Q;U)+D(R:S;V). \text{ -----(8.4)}$$

For all $P,Q\in\Delta_n$, $R,S\in\Delta_m$ and $P*R,\ Q*S\in\Delta_{m}$, $U*V$ is joint utility distribution of U and V.

Bhaker and Hooda [6] also studied generalized mean value characterization of the following measure of 'useful' relative information of $P*P'=(p_1p_1',p_1p_2',\dots\dots\dots p_1p_m',\dots\dots.p_np_1',\dots\dots\dots p_np_m')$ of order α.

$$D_\alpha(P:Q;U) = (\alpha-1)^{-1}\log\frac{\sum_{i=1}^{n} u_i p_i^\alpha q_i^{1-\alpha}}{\sum_{i=1}^{n} u_i p_i}, \ \alpha > 0, \alpha \neq 1 \text{------ (8.5)}$$

The measure (8.5) is additive. So it is also interesting to find out the measure which are non-additive and satisfy the non-additive property of the following form:

$$D(P*Q:Q*S;U*V) = D(P:Q;U) + D(R:S;V) + k\,D(P:Q;U)D(R:S;V), \text{------ (8.6)}$$

Where $P,Q \in \Delta_n, R,S \in \Delta_m$,U and V are utility distribution and $k \neq 0$ is any arbitrary real number.

An another generalization of (6.3) was suggested and studied by Ram (1998) which is given as

$$D^e(P:Q;U) = \frac{1}{\alpha-1}\left[\frac{\sum_{i=1}^{n} u_i p_i^\alpha q_i^{1-\alpha}}{\sum_{i=1}^{n} u_i p_i} - 1\right], \ 0 < \alpha < 1. \text{---(8.7)}$$

Hooda and Kumr [15] proposed a new generalized measure of 'useful' relative information based on (m-1) probability distributions and studied its applications in export industries, ranking and pattern recognition and defense requirements.

Analogously the 'useful' J-divergence corresponding to (8.3) is defined as

$$J(P\,|\,Q;U) = \frac{\sum_{i=1}^{n} u_i p_i \log\frac{p_i}{q_i}}{\sum_{i=1}^{n} u_i p_i} + \frac{\sum_{i=1}^{n} u_i q_i \log\frac{q_i}{p_i}}{\sum_{i=1}^{n} u_i q_i} \text{ -----(8.8)}$$

Most of the measures studied by these authors are additive in nature. The non-additive 'useful' relative information was firstly introduced by Hooda and Tuteja [20] which was defined as follows:

$$D^i(P\,|\,Q;U) = \frac{\sum_{i=1}^{n} u_i p_i \log\left[\left(\frac{p_i}{q_i}\right)^{\beta-1} - 1\right]}{(2^{\beta-1}-1)\sum_{i=1}^{n} p_i}, \ \beta \neq 1. \text{ ---------(8.9)}$$

Discussion and Conclusion

Uncertainty and fuzziness are basic nature of human thinking and of many real objectives. Fuzziness is found in our decision, in our language and in the way we process information. The main use of information is to remove uncertainty which is of two types namely, probabilistic uncertainty and ambiguous or fuzzy uncertainty. The first one was dealt by Shannon [40] defined a new concept entropy which measures the information contained in a probability distribution of a random variable. Thus entropy is information measure that was why many researchers have used information measure in place of entropy.

Zadeh [50] argued that fuzzy uncertainty is different in character from probability uncertainty and introduced fuzzy set by defining a function from a random variable X to [0,1]. Later on analogous to Shannon's entropy, De Luca and Termini [9] defined fuzzy entropy which was also called as fuzzy information measures by many authors. A lot of work have been done in fuzzy mathematics and have studied its application in solving the problems related to the environment of Vagueness. However, there is a scope of further defining new fuzzy information measures and study their applications in solving the problems of decision making, dimension reduction, pattern recognition, medical diagnosis, etc.

Conflict of Interest

It is stated that there is no conflict of interest among authors.

Bibliography

1. Aczel J and Nath P. "Axiomatic characterizations of some measures of divergence in information". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 21.3 (1972): 215-224.
2. Aggarwal NL and Picard C F. "Functional equations and information measures with preference". *Kybernetika* 14.3 (1978): 174-181.
3. Arimoto S. "Information-theoretical considerations on estimation problems". *Information and Control* 19.3 (1971): 181-194.
4. Behara M and Chawla J S. "Generalized gamma-entropy". *Selecta Statistica Canadiana* 2 (1974): 15-38.
5. Belis M and Guiasu S. "A quantitative-qualitative measure of information in cybernetic systems". *IEEE Transactions on Information Theory* 14.4 (1968): 593-594.
6. Bhaker U and Hooda DS. "Mean value characterization of 'useful' information measure". *Tamkang Journal of Mathematics* 24.4 (1993): 383-394.

7. Boekee D E and Van der Lubbe J C. "The R-norm information measure". *Information and Control* 45.2 (1980): 136-155.

8. Burbea J and Rao C R. "Entropy differential metric, distance and divergence measures in probability spaces: A unified approach". *Journal of Multivariate Analysis* 12.4 (1982): 575-596.

9. De Luca A and Termini S. "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory". *Information and Control* 20.4 (1972): 301-312.

10. Emptoz h. "Beta-type information integrating a concept of utility" (1976).

11. Guiasu S. "Information Theory with Applications". McGrew-Hill Mass (1976).

12. Guiasu S and Picard CF. "Brone Inferieure DelaLogueur De Certain Codes". C.R. Academy Science Paris, 273 (1971): 248-251.

13. Hartley R V. "Transmission of information". *Bell System Technical Journal* 7.3 (1928): 535-563.

14. Havrda J and F Charvat. "Quantification of classification processes concept of structural entropy". *Kybernetike* 3 (1967): 30-35.

15. Hooda D S and Kumar P. "Generalised measures of useful directed divergence and information improvement with applications". *Defence Science Journal* 54.2 (2004): 125.

16. Hooda D S and Ram A. "Characterization of generalized R-norm entropy". *Caribbean Journal of Mathematical and Computer Science* 8 (1998): 18-31.

17. Hooda D S and Sharma D K. "Generalized R-norm information measures". *Journal of Applied Mathematics, Statistics and Informatics* 4 (2008): 153-168.

18. Hooda D S and Singh U. "On useful information generating functions". *Statistica* 46 (1986): 528-535.

19. Hooda D S and Tuteja R K. "Two generalized measures of "useful" information". *Information Sciences* 23.1 (1981): 11-24.

20. Hooda D S and Tuteja R K. "On Characterization of non-additive Measures of Relative Information and Inaccuracy". *Bulletin of the Calcutta Mathematical Society* 77 (1985): 363-369.

21. Hooda DS and Kumar A. "The mean and the variance of an MLE of generalized "useful" information". *The Mathematics Student* 49 (1981): 263-269.

22. Jain P and Tuteja R K. "An Axiomatic Characterization of Relative 'Useful'Information". *Journal of Information and Optimization Sciences* 7.1 (1986): 49-57.

23. Jaynes E T. "Information theory and statistical mechanics". *Physical Review* 106.4 (1957): 620.

24. Kapur J N. "Generalized entropy of order α and type β". In The Math Seminar 4 (1967): 78-82).

25. Kapur J N. "On some applications of dynamic programming to information theory". In Proceedings of the Indian Academy of Sciences-Section A 67.1 (1968): 1-11.

26. Kapur J N. "Four families of measures of entropy". *Indian Journal of Pure and Applied Mathematics* 17 (1968a): 429-449.

27. Kapur J N. "New generalized measure of entropy and directed divergence research report". Indian Institute Of Technology, Kanpur 341 (1986b).

28. Kulback S. "Information Theory and statistics". Courier Corporation, USA (1959).

29. Kullback S and Leibler R A. "On information and sufficiency". *The Annals of Mathematical Statistics* 22.1 (1951): 79-86.

30. Kumar P and Hooda DS. "On generalized measures of entropy and dependence, Math". Slovaca (2008).

31. Kumar S., *et al*. "On 'Useful' R-norm Relative Information and J-Divergence Measures". *International Journal of Pure and Applied Mathematics* 77 (2012): 3349-358.

32. Kvålseth TO. "The relative useful information measure: Some comments". *Information Sciences* 56.1-3 (1964): 35-38.

33. Longo G. "Quantitative-qualitative measure of information". New York, NY, USA: Springer (1972).

34. Nyquist H. "Certain factors affecting telegraph speed". *Transactions of the American Institute of Electrical Engineers* 43 (1924): 412-422.

35. Pal N R and Pal S K. "A review on image segmentation techniques". *Pattern Recognition* 26.9 (1993): 1277-1294.

36. Patni G C and Jain K C. "On axiomatic characterization of some non-additive measures of information". *Metrika* 24.1 (1977): 23-34.

37. Prakash O. "Characterization of information measure and information transmission". Ph.d. Thesis, G.N.D. University, Amritsar (India) (1988).

38. Rathie P N and Kannappan P. "A directed-divergence function of type β". *Information and Control* 20.1 (1972): 38-45.

39. Rathie P N and Sheng L T. "The J-divergence of order α". *Journal of Combinatorics, Information and System Sciences* 6 (1981): 197-205.

40. Rényi A. "On measures of entropy and information". In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (1961): 547-561.

41. Shannon C E. "A mathematical theory of communication". *The Bell System Technical Journal* 27.3 (1948): 379-423.

42. Sharma B D and Autar R. "Relative information functions and their type (α, β) generalizations". *Metrika* 21.1 (1974): 41-50.

43. Sharma B D and Mittal D P. "New non-additive measures of relative information". *Journal of Combinatorics Information and System Sciences* 2.4 (1977): 122-132.

44. Sharma B D., *et al*. "On measures of "useful" information". *Information and Control* 39.3 (1978): 323-336.

45. Sharma B D and Taneja I J. "Three generalized-additive measures of entropy". *Elektronische Informationsverarbeitung und Kybernetik* 13.7/8 (1977): 419-433.

46. Taneja H C and Tuteja R K. "Characterization of a quantitative-qualitative measure of relative information". *Information Sciences* 33.3 (1984): 217-222.

47. Taneja I J. "On generalized information measures and their applications". In Advances in Electronics and Electron Physics 76 (1989): 327-413.

48. Theil H. "Economics and information theory (No. 04; HB74. M3, T4.)" (1967).

49. Theil H and Uribe P. "The information approach to the aggregation of input-output tables". *The Review of Economics and Statistics* (1967): 451-462.

50. Zadeh L A. "Electrical engineering at the crossroads". *IEEE Transactions on Education* 8.2 (1965): 30-33.

Volume 3 Issue 10 October 2021