A STUDY OF GENERALIZED INFORMATION MEASURES & THEIR INEQUALITES

Mukesh Kumar Sharma, Dr. Alok Darshan Kothiyal
Research Scholar, SSSUTMS, SEHORE, MP
Research Guide, SSSUTMS, SEHORE, MP

Abstract

A mathematical communication theory in the context of communication theory. It develops a way to research disciplines such as coding theory, semantics, decision theory, economics, radar detection, biology, psychology, and many others that demonstrate the necessity and importance of: information theory. Shannon proposed a probabilistic phenomenon to define communication through the concept of entropy, while at the same time creating the concepts of entropy and mutual information. Entropy is used to measure the uncertainty of a random variable associated with a randomized experiment. On the other hand, the mutual information quantifies the dependency between two random variables. Information theory examines all the theoretical problems associated with the transmission of information via communication channels. This includes studying the measurement of uncertainty (information) and various practical and economical methods for encoding information for transmission. An important feature of Shannon's information theory is that the term information can often have a mathematical meaning as a numerically measurable quantity, based on a probabilistic model, so that solutions to many important memory problems and the transmission of information can be in the form of this measure of information be formulated.

Keyword: Information Measures, Shannon's entropy, Leibler relative-entropy

Introduction

The basic topics of information theory include lossless data compression (eg zip files), lossy data compression (eg mp3 and jpg) and channel coding. The field is at the intersection of mathematics, statistics, computer science, physics, neurobiology and electrical engineering. Their impact has been crucial to the success of space missions, the invention of the CD, the viability of mobile phones, the development of the Internet, the study of linguistics and human perception, understanding of black holes and many other areas. The important subdomains of information theory are source coding, channel coding, algorithmic complexity theory, algorithmic information theory, security of information theory and data analysis. Information. A key measure of information is entropy, which is commonly expressed as the average number of bits needed to store or communicate a symbol in a message. The entropy quantifies the uncertainty in predicting the value of a random variable. For example, specifying the result of a coin toss (two equally probable results) provides less information (lower entropy) than specifying the result of a dice roll (six equally likely outcomes).

Generalized Information Measures

Shannon's entropy was generalized in many ways according to researcher's requirement.

A systematic attempt to develop generalizations of Shannon's entropy was carried out by Rényi [1], who characterized entropy of order α given by

$$H_\alpha(P)=\frac{1}{1-\alpha}\log\sum_{i=1}^{n} p^{\alpha},\ \alpha\neq 1,\ \alpha>0 \qquad (1.1)$$

where α is real parameter. We can easily verify that Eq. (1.1) reduces to Shannon entropy defined by Eq. (1.1) as α →1.

$$H_{\alpha,\beta}(P)=\frac{1}{\alpha-\beta}\log\left(\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\beta}}\right),\quad \alpha\neq\beta,\alpha>0,\ \beta>0 \qquad (1.2)$$

where α and β are real parameters

$$H^{\alpha}(P)=(2^{1-\alpha}-1)^{-1}\left[\sum_{i=1}^{n} p_i^{\alpha}-1\right],\quad \alpha\neq 1,\ \alpha>0 \qquad (1.3)$$

for all $P=(p_1,p_2,\ldots\ldots p_n)\in\Delta_n$. In this case, we can also verify that Eq (1.3) reduces to Shannon entropy defined by Eq. (1.1) as1→α.

Sharma and Taneja studied a generalization of Eq. (1.3) involving two scalar parameters, known entropy of degree, (α, β) and is given by

$$H^{\alpha,\beta}(P)=(2^{1-\alpha}-2^{1-\beta})^{-1}\sum_{i=1}^{n}(p_i^{\alpha}-p_i^{\beta}),\quad \alpha\neq\beta,\quad \alpha,\ \beta>0 \qquad (1.4)$$

for all $P=(p_1,p_2,\ldots\ldots p_n)\in\Delta_n$, where α and β are real parameters.

In particular, when $\alpha=1$ or $\beta=1$, the measure given by Eq. (1.3) reduces to Eq. (1. 4). In the limiting case, we have

$$\lim_{\alpha\to\beta} H^{\alpha,\beta}(P)=-2^{\alpha-1}\sum_{i=1}^{n} p_i^{\alpha}\log p_i,\quad \alpha>0$$

it reduces to Shannon entropy for 1→α.

$$H_t(P) = (2^{t-1} - 1)\left[\left(\sum_{i=1}^{n} p_i^{1/t}\right)^t - 1\right], \; t \neq 1, \; t > 0 \qquad (1.5)$$

for all $P = (p_1, p_2, \ldots\ldots p_n) \in \Delta_n$. In this case also we can easily verify that Eq (1.5) reduces to Shannon entropy as t → 1.

Sharma and Mittal [1] introduced and characterized two entropies called entropy of order 1and degree β and entropy of order α and degree β given by

$$H_1^\beta = (2^{1-\beta} - 1)^{-1}\left[\exp_2\left((\beta - 1)\sum_{i=1}^{n} p_i \log p_i\right) - 1\right], \; \beta \neq 1 \qquad (1/6)$$

$$H_\alpha^\beta = (2^{1-\beta} - 1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1\right], \; \alpha \neq 1, \; \beta \neq 1, \; \alpha > 0 \qquad (1.7)$$

respectively, for all $P = (p_1, p_2, \ldots\ldots p_n) \in \Delta_n$ where α and β are real parameters.

Sharma and Mittal's main motivation was to generalize the three entropies, $H\alpha(P)$, $H^\beta(P)$ and $H_t(P)$. With this aim, they arrived at $H^\beta{}_\alpha(P)$. The measure $H^\beta{}_\alpha(P)$ reduces to $H^\beta{}_\alpha(P)$ reduces to $H^\beta{}_1(P)$ and $H_\alpha(P)$, when $\alpha \to 1$ and $\beta \to 1$ respectively. Also, $H^\beta{}_1(P)$ reduces to Shannon's entropy, $H(P)$, when $\beta \to 1$.

Thus, we see that the entropy of order α and degree β contain, either as a limiting or as a particular case, the Shannon's entropy, the entropy of order α, the entropy of degree β, the entropy of kind t, and the entropy of order 1 and degree β.

Relative Information and Inaccuracy

Kullback and Leibler's measure of information associated with the probability distributions P and Q is given by

$$D(P\|Q) = \sum_{i=0}^{n} p_I \log (p_i/q_i) \qquad (1.8)$$

The measure given by Eq. (1.9) has many names given by different authors such as, relative information, directed divergence, cross entropy, function ofdiscrimination etc. Here we shall refer it relative information. It has found many applications in setting important theorems in Information theory and Statistics.

The Kerridge's measure of information generally referred asinaccuracy associated with two probability distributions is given by

$$H(P\|Q) = \sum_{i=0}^{n} p_I \log q_i \qquad (1.9)$$

Divergence Measures

We see that the measure given by Eq. (1. 8) is not symmetric in P and Q. Its symmetric version known as J-divergence (Jeffreys [11], Kullback and Leibler [12]) is given by

$$J(P\|Q) = D(P\|Q) + D(Q\|P) = \sum_{i=0}^{n}(p_i - q_i) \log \left(\frac{p}{q}\right) \qquad (1.10)$$

Sibson for the first time introduced the idea of information radius for ameasure arising due to concavity property of Shannon's entropy. This measure referred as Jensen difference divergence measure is given by

$$I(P \| Q) = H\left(\frac{P+Q}{2}\right) - \frac{H(P)+H(Q)}{2}$$

$$= \sum_{i=1}^{n}\left[\frac{p_i \log p_i + q_i \log q_i}{2} - \left(\frac{p_i + q_i}{2}\right)\log\left(\frac{p_i + q_i}{2}\right)\right] \qquad (1.11)$$

By simple calculations, one can also write

$$I(P \| Q) = \frac{1}{2}\left[D\left(P \| \frac{P+Q}{2}\right) + D\left(Q \| \frac{P+Q}{2}\right)\right] \qquad (1.12)$$

Taneja studied an another kind of measure based on arithmetic and geometric mean inequality calling arithmetic and geometric mean divergence measure given by

$$T(P\|Q)=\frac{1}{2}\left[D\left(\frac{P+Q}{2}\|P\right)+D\left(\frac{P+Q}{2}\|Q\right)\right]$$

$$=\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2\sqrt{p_iq_i}}\right). \quad (1.13)$$

Interestingly these three measures satisfy the following inequality

$$I(P\|Q) + T(P\|Q) = 4J(P\|Q)$$

$H(P;U) = -\sum_{i=0}^{n} u_i p_i \log p_i$, $p_i \geq 0$, $\sum_{i=0}^{n} p_i = 1$, $u_i > 0$ (1.14)

Sharma-Mohan-Mitter also considered utility of the event and studiedindependently Eq. (1. 14) and called it 'Useful' measure of information and generalized Eq. (1.14) 'Useful' information/entropy of type β as follows:

$$H^{\beta}(P;U)=\frac{\sum_{i=1}^{n}u_ip_i(p_i^{\beta-1}-1)}{\sum_{i=1}^{n}u_ip_i},\quad \beta\neq 1,\ \beta>0 \quad (1.15)$$

$$I_M^{\beta}(P;V)=\frac{1}{2^{1-\beta}-1}\left[2^{(1-\beta)\left(\sum_{i=1}^{n}v_i\log\frac{1}{p_i}\right)/\sum_{i=1}^{n}v_i}-1\right],\quad \beta\neq 1,\ \beta>0 \quad (1.16)$$

$$I_M^{\alpha,\beta}(P;V)=\frac{1}{2^{1-\beta}-1}\left[\left\{\frac{\sum_{i=1}^{n}p_i^{\alpha-1}v_i}{\sum_{i=1}^{n}v_i}\right\}^{\frac{\beta-1}{\alpha-1}}-1\right],\quad \alpha\neq 1,\beta\neq 1,\ \alpha\neq\beta,\ \alpha,\beta>0 \quad (1.17)$$

$$H_1^{\alpha}(P;U)=-2^{1-\alpha}\sum_{i=1}^{n}u_ip_i\log p_i,\quad \alpha\neq 1,\ \alpha>0 \quad (1.18)$$

$$H^{\alpha,\beta}(P;U)=(2^{1-\alpha}-2^{1-\beta})^{-1}\sum_{i=1}^{n}u_i(p_i^{\alpha}-p_i^{\beta}) \tag{1.19}$$

$$H_{\alpha}{}^{\beta}(P;U)=(2^{1-\alpha}-2^{1-\beta})^{-1}\frac{\sum_{i=1}^{n}u_i^{\alpha}p_i^{\alpha}(p_i^{\beta-\alpha}-1)}{\sum_{i=1}^{n}p_i},\ \alpha\neq\beta \tag{1.20}$$

Inequalities

The very first inequality known as Shannon's Inequality

$$-\sum p_i\log q_i\geq-\sum p_i\log p_i$$

has played a vital role in coding theory. The inequality on right is the Shannon's entropy and on the left hand side, it is Kerridge inaccuracy. It has been generalized in terms of functions such as

$$\sum p_i f(q_i)\leq\sum p_i f(p_i)$$

$$f(P)=a\log p+b$$

$$\sum_{i=1}^{n}p_iq_i^{w}\begin{cases}\leq\left(\sum_{i=1}^{n}p_iq_i\right)^{w}, & 0<w<1\\ \geq\left(\sum_{i=1}^{n}p_iq_i\right)^{w}, & w>1\ or\ \ w<0\end{cases}$$

with equality if for some $c,\quad q_i^{w}=cb_i^{w/w-1},\ \forall i$ where q_i and b_i are non-negative real numbers, for $w<0$, $q_i>0$, $b_i>0$, $\forall_i$.

$$\left(\sum_{i=1}^{n} p_i^{w}\right)^{1/w}\left(\sum_{i=1}^{n} p_i^{\frac{w}{w-1}}\right)^{\frac{w-1}{w}}\begin{bmatrix} \leq \sum_{i=1}^{n} p_i q_i, & w<1, \quad w\neq 0 \\ \\ \geq \sum_{i=1}^{n} p_i q_i, & w>1 \end{bmatrix}$$

We generalize these information theoretic inequalities that are based on the Gaussian setting to generic ones in the stable setting which coincide with the regular identities in the Gaussian setup.

Conclusion

Information theory based on Shannon entropy functional found applications that cut across a myriad of fields, because of its established *mathematical significance* i.e., its beautiful mathematical properties. Shannon (1956) too emphasized that “the hard core of information theory is, essentially, a branch of mathematics” and “a thorough understanding of the mathematical foundation is surely a prerequisite to other applications.” Given that “the hard core of information theory is a branch of mathematics,” one could expect formal generalizations of information measures taking place, just as would be the case for any other mathematical concept. Generalized information measures to the measure-theoretic case. We showed that as in the case of Kullback- Leibler relative-entropy, generalized relative-entropies, whether R´enyi or Tsallis, in the discrete case can be naturally extended to measure-theoretic case, in the sense that measure-theoretic definitions can be derived from limits of sequences of finite discrete entropies of pmfs which approximate the pdfs involved. We also showed that ME prescriptions of measure-theoretic Tsallis entropy are consistent with the discrete case, which is also true for measure-theoretic Shannon-entropy.

References

[1] Aczel, J. (1966), “Lectures on functional equations and their applications”, Academic Press, New York, vol. 19.

[2] Aczel, J. and Darcozy, Z. (1975), “On measures of information and their characterization”, Academic Press, New York, vol. 19.

[3] Behra, M. (1981), “Additive and non additive measure of entropy”, Wiley Eastern, New Delhi.

[4] Behra, M. and Chawla, J. S. (1974),“Generalized gamma entropy”, Selecta Statistica Canadiana, vol. 2, no. 1-2, pp. 15-38.

[5] Chaundy, T. W. and McLeod, J. B. (1960), “On a functional equation”, Quaterly Journal of Mathematics, vol. 9, no. 2, pp. 202-206.

[6] Darcozy, Z. (1970), "Generalized information function", Information and Control, vol. 16, no. 1, pp. 36-51.

[7] Duncan, G. T. (1974), "Heterogeneous questionnaire theory", Society for Industrial and Applied Mathematics, Journal of Applied Mathematics, vol. 27, no. 1, pp. 59-71.

[8] Hartley, R. T. V. (1928), "Transmission of information", Bells System Technical Journal, Vol. 7, no. 3, pp. 535-563.

[9] Havrda, J. F. and Charvat, F. (1967), "Quantification method of classification Processes, the concept of structural α –entropy", Kybernetika, vol. 3, no. 1, pp. 30-35.

[10] Hooda, D. S. (1986), "On a generalized measure of relative useful information", Soochow Journal of Mathematics, vol. 12, no. 1, pp. 23-32.

[11] Jain P., Taneja, H. C. and Tuteja, R. K. (1986), "Sub-additive measures of relative 'useful' information", Tamkang Journal of Mathematics, vol. 17, no. 3, pp. 97-104.

[12] Jain, P. and Tuteja, R. K. (1989), "On coding theorems connected with 'useful' entropy of order β", International Journal of Mathematics and Mathematical Sciences, vol. 12, no. 1, pp. 193-198.