A Theoretical Study of Renyi's Measures of Entropy

Deepak Kumar and Vishal Rana

Department of Mathematics, School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara, Punjab, India.

Abstract

The present manuscript deals with the study of various entropy measures. The entropy measures can be broadly categorized into two sections namely additive measures and non-additive measures. The present manuscript is divided into three sections. In the first section, introduction of entropy measures and definition of entropy is given. The second section deals with the various requirements of measures of entropy. In the third section, the existing additive measures of entropy known as **Renyi's measure** has been studied and all the requirements for the measures have been verified

1) Introduction

Information theory was created by C. Shannon in 1948 so as to address the hypothetical inquiries in media communications. In information theory, entropy is a measure of the arbitrariness of a discrete random variable. It can likewise be thought of as the uncertainty about the result of an experiment, or the rate of information generation by playing out the experiment repeatedly. The idea of entropy was acquainted with giving a quantitative measure of uncertainty.

Shannon [1] determined the measure $H(P) = -\sum_{i=1}^{n} p_i \, ln \, p_i$ for the uncertainty of a probability distribution $(p_1, p_2, \dots p_n)$ and defined it as entropy. The information theoretic entropy can be estimated as far as its error from the uniform distribution which is the unsure distribution. Following the Shannon's measure of entropy, countless measures of information theoretic entropies have been determined. Renyi [2] described entropy of order α as $H_\alpha(P) = \frac{1}{1-\alpha}\left[\sum_{i=1}^{n} p_i^\alpha \Big/ \sum_{i=1}^{n} p_i\right], \alpha \neq 1, \alpha > 0$, which speaks to a group of measures which incorporates Shannon's entropy as a restrictive case as α→1. Later, Kapur [3] summed up Renyi's measure further to give a measure of entropy of order 'α' and type 'β', *viz.*,

$$H_{\alpha,\beta}(P) = \frac{1}{1-\alpha} \ln\left[\sum_{1=1}^{n} p_i^{\alpha+\beta-1} \Big/ \sum_{i=1}^{n} p_i^\beta\right], \alpha \neq 1, \alpha > 0, \beta > 0, \alpha + \beta - 1 > 1,$$

This decreases to Renyi's measure when β=1, to Shannon measure, when$\beta = 1, \alpha \to 1$. When$\beta = 1, \alpha \to \infty$, it gives the measure $H_\infty(P) = -ln \, P_{max}$.

Havrada and Charvat [4] introduced the first non-additive measure of entropy specified by

$$H^{\alpha}(P) = \frac{[\sum_{i=1}^{n} p_i^{\alpha}]-1}{2^{1-\alpha}-1}, \alpha \neq 1, \alpha > 0.$$

To be predictable with Renyi's measure and for numerical comfort, it is utilized in changed structure as

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

Behara and Chawla [5] characterized the non additive γ-entropy as

$$H_{\gamma}(P) = \frac{1-\left(\sum_{i=1}^{n} p_i^{1/\gamma}\right)}{1-2^{\gamma-1}}, \gamma > 0, \gamma \neq 0$$

$$= \frac{1}{1-2^{\gamma-1}} - \frac{1}{1-2^{\gamma-1}}\left[\sum_{i=1}^{n} p_i^{1/\gamma}\right]^{\gamma}$$

Definition 1:_The entropy is defined as lack of order or predictability, gradual decline into disorder. In thermodynamics, it is characterized as the thermodynamic amount speaking to the inaccessibility of a system's thermal energy for transformation into mechanical work, frequently translated as the level of confusion or arbitrariness in the system.

Examples: Ice softening, salt or sugar dissolving, making popcorn and bubbling water for tea are process with expanding entropy.

1.1) Requirements of Measure of Entropy

Let the probabilities of an possible outcomes $A_1, A_2, \dots A_n$ of an experiment be respectively $p_1, p_2 \dots p_n$ offering ascend to the probability distribution $P = (p_1, p_2 \dots p_n)$;

$$\sum_{i=1}^{n} p_i = 1,\ p_1 \geq 0, p_2 \geq 0, \dots p_n \geq 0$$

There is uncertainty with regards to the result when the experiment is done. Any measure of this uncertainty should satisfy the following requirements:

1) It ought be a function of $p_1, p_2, \dots p_n$, so that we may write down it as
$$H(P) = H_n(p) = H_n(p_1, p_2, \dots p_n)$$
2) It ought be uniform function of $p_1, p_2, \dots p_n$ i.e. little change in $p_1, p_2, \dots p_n$ should cause a little change in H_n.
3) It ought not alter when the outcomes are rearranged among themselves .i.e. H_n ought to be ordered function of its contentions.
4) It ought not change if an unthinkable result is added to the probability scheme i.e.

$$H_{n+1}(p_1, p_2, \dots p_n, 0) = H_n(p_1, p_2, \dots p_n)$$

5) It ought be minimum and possibly zero at the point when there is no uncertainty about the result. Along these lines, it ought to disappear when one of the results is sure to occur so that $H_n(p_1, p_2, \dots p_n) = 0, \sum_{i=1}^{n} p_i = 1, \sum_{j=1}^{m} p_j = 1,\ j \neq i, i = 1,2, \dots n$
6) It ought to be greatest when there is a most extreme uncertainty which rises when the results are similarly likely so that H_n should be maximum when $p_1 = p_2 = \cdots = p_n = {}^1\!/_n$.
7) The greatest estimation of H_n should increment as n increments.
8) For two self-determining probability distribution

$$P = (p_1, p_2, \dots p_n), Q = (q_1, q_2, \dots q_n), \sum_{i=1}^{n} p_i = 1, \sum_{j=1}^{m} q_j = 1$$

The uncertainty of the combined scheme $P \cup Q$ ought to be their addition of their vulnerabilities i.e. $H_{nm}(P \cup Q) = H_n(P) + H_m(Q)$, where if $A_1, A_2, \dots A_n;\ B_1, B_2, \dots B_n$ are the outcomes of P and Q then the outcomes of $P \cup Q$ are A_i, B_j with probabilities $p_i q_j (i = 1,2, \dots n, j = 1,2, \dots m)$.

2) Main Section

In this section, we have presented a discussion on the existing additive measure of entropy called **Renyi's Measure of Entropy** and verified all the requirements for the existing measure:

Renyi [2] suggested the following measure of entropy:

$$R(P)\ or\ H_\alpha(P) = \frac{1}{1-\alpha} ln \frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i}, \qquad –(1)$$

which is called Renyi's entropy of order α, in this part, we study some of the properties of measure of uncertainty and conclude how these measures make Renyi's measure a most satisfactory measure of entropy.

(i) Equation (1) is a function of $p_1, p_2, \dots p_n$.

(ii) Consider $H_\alpha(P) = \frac{1}{1-\alpha} ln \frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i}$

$$= \frac{1}{1-\alpha} ln \sum_{i=1}^{n} p_i^\alpha$$

Here, $ln \sum_{i=1}^{n} p_i^\alpha$ is a uniform function of $p_1, p_2, \dots p_n$ and little change in $p_1, p_2, \dots p_n$ cause little change in $H_\alpha(P)$.

(iii) $H_\alpha(P)$ is permutationally uniform, it does not alter if $p_1, p_2, \dots p_n$ are reordered amongst themselves.

(iv) The entropy does not alter by the addition of a not possible event i.e. of an event with zero possibility, thus,

$$H_\alpha(p_1,p_2,\ldots,p_n,0)=\frac{1}{1-\alpha}\ln[\sum_{i=1}^{n}p_i^\alpha+0^\alpha]$$

$$=\frac{1}{1-\alpha}ln\sum_{i=1}^{n}p_i^\alpha$$

Therefore, $H_\alpha(p_1,p_2,\ldots,p_n,0)=\mathrm{H}_\alpha(p_1,p_2,\ldots,p_n)$

(v) There are n degenerate distributions

$\varDelta_1=(1,0,0,\ldots 0)$

$\varDelta_2=(0,1,0,\ldots 0)$

…

…

…

$\varDelta_n=(0,0,0,\ldots 1)$

For every one of these $H_\alpha(P)=0$,we imagine that for every one of these distribution, the uncertainty should be zero. Renyi satisfies this condition as $ln(1^\alpha)=ln(1)=0$.

(vi) We use Lagrange's way to raise the entropy subjected to$\sum_{i=1}^{n}p_i=1$, In this case Lagrangian is

$$L=\frac{1}{1-\alpha}\ ln\frac{\sum_{i=1}^{n}p_i^\alpha}{\sum_{i=1}^{n}p_i}+\lambda(\sum_{i=1}^{n}p_i-1)$$

$$L=\frac{1}{1-\alpha}\ ln\sum_{i=1}^{n}p_i^\alpha+\lambda(\sum_{i=1}^{n}p_i-1) \quad –(2)$$

Differentiating equation (2) partially $w.r.t.p_1,p_2,\ldots p_n$, we get

$$\frac{\partial L}{\partial p_1}=\frac{1}{1-\alpha}\frac{1}{\sum_{i=1}^{n}p_i^\alpha}\alpha p_1^{\alpha-1}+\lambda$$

$$\frac{\partial L}{\partial p_2}=\frac{1}{1-\alpha}\frac{1}{\sum_{i=1}^{n}p_i^\alpha}\alpha p_2^{\alpha-1}+\lambda$$

…

…

…

$$\frac{\partial L}{\partial n}=\frac{1}{1-\alpha}\frac{1}{\sum_{i=1}^{n}p_i^\alpha}\alpha p_n^{\alpha-1}+\lambda$$

Equating $\frac{\partial L}{\partial p_1}, \frac{\partial L}{\partial p_2}, \ldots \frac{\partial L}{\partial p_n}$ equal to zero, we get

$$\frac{1}{1-\alpha}\frac{\alpha p_1^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\alpha}} = \frac{1}{1-\alpha}\frac{\alpha p_2^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\alpha}} = \cdots = \frac{1}{1-\alpha}\frac{\alpha p_n^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\alpha}}$$

⇨ $p_1^{\alpha-1} = p_2^{\alpha-1} = \cdots = p_n^{\alpha-1}$

⇨ $p_1 = p_2 = \cdots = p_n$

But $\sum_{i=1}^{n} p_i = 1$

⇨ $p_1 + p_2 + \cdots + p_n = 1$

⇨ $np_1 = 1[\because p_1 = p_2 = \cdots = p_n]$

⇨ $p_1 = \frac{1}{n}$

⇨ Thus, $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$,

and $\lambda = \frac{\alpha}{1-\alpha}\frac{(\frac{1}{n})^{\alpha-1}}{\sum_{i=1}^{n}(\frac{1}{n})^{\alpha}}$

The 2nd order Hessian matrix is

$$\begin{bmatrix} \frac{\partial^2 L}{\partial p_1^2} & \frac{\partial^2 L}{\partial p_2 p_1} & \cdots & \frac{\partial^2 L}{\partial p_n p_1} \\ \frac{\partial^2 L}{\partial p_1 p_2} & \frac{\partial^2 L}{\partial p_2^2} & & \frac{\partial^2 L}{\partial p_n p_2} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 L}{\partial p_1 \partial p_n} & \frac{\partial^2 L}{\partial p_2 \partial p_n} & \cdots & \frac{\partial^2 L}{\partial p_n^2} \end{bmatrix}$$

To prove the condition of maxima, consider $P = (p_1, p_2)$ and assume that $\alpha = 2 \neq 1$ and $\alpha > 0$.

Thus, the 2nd order Hessian matrix for case reduces to

$$\begin{bmatrix} \frac{\partial^2 L}{\partial p_1^2} & \frac{\partial^2 L}{\partial p_1 \partial p_2} \\ \frac{\partial^2 L}{\partial p_1 \partial p_2} & \frac{\partial^2 L}{\partial p_2^2} \end{bmatrix}$$

and value of L becomes $-\ ln\sum_{i=1}^{2} p_i^2 + \lambda(\sum_{i=1}^{2} p_i - 1)$

i.e. $L = -ln\sum_{i=1}^{2} p_i^2 + \lambda(\sum_{i=1}^{2} p_i - 1)$ –(3)

where $\Psi = p_1 + p_2 - 1$

Differentiating equation (3) partially $w.r.t.\ p_1, p_2$, we get

$$\frac{\partial L}{\partial p_1} = \frac{-1}{\sum_{i=1}^{2} p_i^2}.2p_1 + \lambda$$

$$\frac{\partial L}{\partial p_2} = \frac{-1}{\sum_{i=1}^{2} p_i^2}.2p_2 + \lambda \ and \ \frac{\partial L}{\partial \lambda} = p_1 + p_2 - 1$$

Equating $\frac{\partial L}{\partial p_1}, \frac{\partial L}{\partial p_2}, \frac{\partial L}{\partial \lambda}$ equal to zero, we get

$$\frac{\partial L}{\partial p_1} = 0 \ => \frac{2p_1}{\sum_{i=1}^{2} p_i^2} = \lambda$$

$$\frac{\partial L}{\partial p_2} = 0 \ => \frac{2p_2}{\sum_{i=1}^{2} p_i^2} = \lambda$$

$$\frac{\partial L}{\partial \lambda} = p_1 + p_2 - 1 = 0$$

⇨ $p_1 = 1 - p_2$

Thus, $\frac{2p_1}{\sum_{i=1}^{2} p_i^2} = \frac{2p_2}{\sum_{i=1}^{2} p_i^2}$

⇨ $p_1 = p_2$

But $p_1 + p_1 = 1$

$2p_1 = 1$

⇨ $p_1 = \frac{1}{2}$

Thus, $p_1 = p_2 = \frac{1}{2}$ and $\lambda = 2$

Differentiating $\frac{\partial L}{\partial p_1}$ $w.r.t. p_1 \ and \ p_2$, we get

$$\frac{\partial^2 L}{\partial p_1^2} = -2[\frac{\sum_{i=1}^{2} p_i^2 1 - p_1 2p_1}{(p_1^2 + p_2^2)^2}]$$

$$= -2[\frac{p_1^2 + p_2^2 - 2p_1^2}{(p_1^2 + p_2^2)^2}]$$

$$= -2[\frac{p_2^2 - p_1^2}{(p_1^2 + p_2^2)^2}]$$

$$= 2[\frac{(p_1^2 - p_2^2)}{(p_1^2 + p_2^2)^2}]$$

and $\frac{\partial^2 L}{\partial p_2 \partial p_1} = \frac{4p_1 p_2}{(p_1^2 + p_2^2)^2}$

Differentiating $\frac{\partial L}{\partial p_2}$ $w.r.t. p_2$ and p_1, we get

$$\frac{\partial^2 L}{\partial p_2^2} = -2[\frac{\sum_{i=1}^{2} p_i^2 - p_2 2p_2}{(p_1^2 + p_2^2)^2}]$$

$$= -2[\frac{p_1^2 + p_2^2 - 2p_2^2}{(p_1^2 + p_2^2)^2}]$$

$$= -2[\frac{p_1^2 - p_2^2}{(p_1^2 + p_2^2)^2}]$$

$$= 2\left[\frac{(p_2^2 - p_1^2)}{(p_1^2 + p_2^2)^2}\right],$$

and $\frac{\partial^2 L}{\partial p_1 \partial p_2} = \frac{4p_1 p_2}{(p_1^2 + p_2^2)^2}$

Now, $\frac{\partial^2 L}{\partial p_1^2} = 0$ at $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$

$\frac{\partial^2 L}{\partial p_2^2} = 0$ at $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$

$\frac{\partial^2 L}{\partial p_1 \partial p_2} = \frac{\partial^2 L}{\partial p_2 \partial p_1} = 4$

$\frac{\partial \Psi}{\partial p_1} = 1$ at $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$

$\frac{\partial \Psi}{\partial p_2} = 1$ at $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$

2nd order condition is

$$|H_2| = \begin{vmatrix} 0 & \Psi_1 & \Psi_2 \\ \Psi_1 & L_{11} & L_{12} \\ \Psi_2 & L_{21} & L_{22} \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 4 \\ 1 & 4 & 0 \end{vmatrix} = 8 > 0$$

So, L has maximum value at $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$, on generalizing it, we get

$L = \frac{1}{1-\alpha} ln \frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i} + \lambda(\sum_{i=1}^{n} p_i - 1)$ has maximum value at $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$.

(vii) The maximum value of H_α is given by

$$H_\alpha(P) = \frac{1}{1-\alpha} ln \sum_{i=1}^{n} (\frac{1}{n})^\alpha$$

$$= \frac{1}{1-\alpha} \ln n \left(\frac{1}{n}\right)^{\alpha}$$

$$= \frac{1}{1-\alpha} ln\, n^{1-\alpha}$$

$$= \frac{1}{1-\alpha} (1-\alpha) ln\, n$$

$$= \ln n$$

Since, $\ln n$ is an rising function of n, so is $H_n(\alpha)$. Thus, there should be a rise in maximum uncertainty when more outcomes are possible.

(viii) Let $P = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_m)$ be two self-determining probability distribution of two random variables X and Y so that $P(X = x_1) = p_1, P(Y = y_i) = p_i q_j$.

For the combined distribution of X and Y, there are mn possible outcomes, with probability $p_i q_j$ for i$= 1,2, \ldots n$ and $j = 1,2, \ldots m$ so that for the combined probability distribution which we shall now denote by $p * q$, the entropy is given by

$$H_{mn}(p * q) = \frac{\alpha}{1-\alpha} ln \sum_{i=1}^{n} \sum_{j=1}^{m} (p_i q_j)^{\alpha}$$

$$= \frac{\alpha}{1-\alpha} ln \sum_{i=1}^{n} p_i^{\alpha} \sum_{j=1}^{m} q_j^{\alpha}$$

$$= \frac{\alpha}{1-\alpha} \{ln \sum_{i=1}^{n} p_i^{\alpha} + ln \sum_{j=1}^{m} q_j^{\alpha}\}$$

$$= \frac{\alpha}{1-\alpha} \quad ln \sum_{i=1}^{n} p_i^{\alpha} + \frac{\alpha}{1-\alpha} ln \sum_{j=1}^{m} q_j^{\alpha}$$

$$= H_m(P) + H_m(q)$$

For two self-determining distributions, the entropy of the combined distribution is the addition of the entropies of the two distributions, which is the desirable property and this is called the additive property of the measure of entropy.

Remark:-The Renyi's entropy of order α is specified by

$$R(P) = \frac{1}{1-\alpha} ln \frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i}, \alpha \neq 1, \alpha > 0$$

$$= \frac{1}{1-\alpha} ln \sum_{i=1}^{n} p_i^{\alpha}$$

Therefore, $\lim_{\alpha \to 1} R(P) = \lim_{\alpha \to 1} \frac{1}{1-\alpha} ln \sum_{i=1}^{n} p_i^{\alpha} \qquad [\frac{0}{0} form]$

$$= \lim_{\alpha \to 1} \frac{\frac{1}{\sum_{i=1}^{n} p_i^{\alpha}} \sum_{i=1}^{n} p_i^{\alpha} . ln p_i^{\alpha}}{-1},$$

which is the Shannon measure of entropy. Hence, Shannon measure of entropy is the restrictive case of Renyi's measure of entropy.

3) CONCLUSION

In the present manuscript, we have verified the various requirements for the existing additive measure of entropy and studied about the Renyi's measure of entropy to answer the theoretical questions in telecommunications. The concept was of transferring maximum information through a noisy channel with negligible error. But there were some limitations of his theory. Thus, many researchers gave their measures to increase the efficiency of transferring information with minimized loss of energy and reducing the error rate of data.

REFERENCES

[1] Shannon, C.E. (1948): "A mathematical theory of communication", Bell. Sys. Tech. Jr., 27, 379-423, 623-659.

[2] Renyi, A. (1961): "On measures of entropy and information", Proc. 4th Ber. Symp. Math. Stat. and Prob., 1, 547-561.

[3] Kapur, J.N. (1967): "Generalized entropy of order α and type β", Maths. Semi. 4, 79-84.

[4] Havrada, J.H. and Charvat, F. (1967): "Quantification methods of classification processes: Concept of structural α-entropy", Kybernetika, 3, 30-35.

[5] Behara, M. and Chawla, J.S. (1974): "Generalized γ-entropy", Selecta Statistica Canadiana, 2, 15-38.