RESEARCH REVIEW JOURNALS
Volume-04
Issue-01
January -2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Characterization of New Information Measures Via Measures of Central Tendency and Dispersion

[1]Vishal & [2]Prachi

[1,2]*Department of Mathematics, S.S.M. College, Dinanagar (India)*

ARTICLE DETAILS

Article History
Published Online: 20 January 2019

Keywords
Entropy; concavity; symmetry; standard deviation; arithmetic mean; harmonic mean.

ABSTRACT
It is known fact that statistical constants play an extensive role in the field of Statistics. On the other hand, measures of information are used to study uncertainty, diversity, randomness etc. present in different phenomenon. These two fields which are used independent of each other can be clubbed together. In this communication, we have developed the link between the two and proved that the well known standard statistical measures like measures of central tendency can be used as information theoretic measures.

1. Introduction

It is well known that the most successful measure of information is certainly Shannon's [6] entropy, which is interpreted as the uncertainty content of a random experiment ruled by the probability distribution P . Shannon [6] introduced the concept of information theoretic entropy associated with every probability distribution $P = (p_1 , p_2 , ..., p_n)$ where $p_1, p_2 , ..., p_n$, are the probabilities of n outcomes. He laid down certain very plausible postulates which this measure should possess and found that the only function which satisfies these postulates is given by

$$H(P) = -\sum_{i=1}^{n} p_i \operatorname{In} p_i \qquad (1.1)$$

After Shannon's [6] measure of entropy, Renyi [5] introduced entropy of order α, given by the following expression:

$$H_a(P) = \frac{1}{1-\alpha} \operatorname{In} \left[\sum_{i=1}^{n} p_i^a \Big/ \sum_{i=1}^{n} p_i \right], a \neq 1, a > 0 \qquad (1.2)$$

Havrada and Charvat [2] introduced another measure called the non-additive measure of entropy, given by:

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

Burg [1] developed his non-parametric measure of entropy, given by

$$H^1(P) = \sum_{i=1}^{n} \log p_i \qquad (1.4)$$

Many other probabilistic measures of entropy have been discussed by Kapur [4], Herremoes [3], Sharma and Taneja [7], Zadeh [10] etc.

In Biological systems, there exist many well-known measures of information which are frequently used for measuring diversity in different communities. Some of these measures are due to Shannon [6], Renyi [5], Simpson [8], Weiner [9] etc. Of course Shannon's [6] measure is most widely applicable and possesses many interesting properties. But, it has a limitation that it deals with exponential families only. In actual practice, there are distributions which are non-exponential. Thus, there is a need for developing new measures to extend the scope of their applications. Such measures have been developed in this paper.

2. New Information Measures Via Measures of Central Tendency and Dispersion

In this section, we develop the following information theoretic measures:

(I) Information Measure in terms of Arithmetic Mean, Harmonic Mean and Standard Deviation

Let a random variable X takes values $x_1, x_2, x_3, ..., x_n$. Then, geometric mean G, arithmetic mean M and harmonic mean H of these n observations are given by:

$$G = (x_1 . x_2 . x_3 . x_n)^{\frac{1}{n}} \quad x_i \geq 0 \tag{2.1}$$

$$M = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.2}$$

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + .. + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{2.3}$$

Equations (2.1), (2.2) and (2.3) can be rewritten as

$$G = (p_1 . p_2 . p_3 .... p_n)^{\frac{1}{n}} \cdot \sum_{i=1}^{n} x_i \tag{2.4}$$

$$\text{where } p_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

$$\text{From (2.2), } nM = \sum_{i=1}^{n} x_i \tag{2.5}$$

$$\text{From (2.2) and (2.3), } \frac{H}{M} = \frac{n^2}{\sum_{i=1}^{n} \frac{1}{p_i}} \tag{2.6}$$

We know that the variance of discrete distribution of n observations $x_1, x_2, x_3, ..., x_n$ is defined as **(2.7)**

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]^2$$

$$= \frac{1}{n}(p_1^2 + p_2^2 + \ldots + p_n^2)\left(\sum_{i=1}^{n} x_i\right)^2 - \frac{1}{n2}\left(\sum_{i=1}^{n} x_i\right)^2$$

This can be rewritten as

$$\sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{n} p_i^2 - \frac{1}{n}\right](nM)^2$$

Thus, we have (2.8)

$$\frac{\sigma^2 + M^2}{nM^2} = \left(\sum_{i=1}^{n} p_i^2\right)$$

With simple calculations, equations (2.6) and (2.8) give the following result:

$$\text{Log}\,\frac{H}{n^2 M} - \frac{\sigma^2 + M^2}{nM^2} = -\log\left\{\sum_{i=1}^{n}\frac{1}{p_i^2}\right\} - \sum_{i=1}^{n} p_i^2 \tag{2.9}$$

We shall prove that the R.H.S.of (2.9) is an information measure. Thus, we consider the following function

$$\psi^n_n(P) = -\log\left\{\sum_{i=1}^{n}\frac{1}{p_i^2}\right\} - \sum_{i=1}^{n} p_i^2 \tag{2.10}$$

We have

$$\psi^n_n(P) = \frac{1}{p_i^4\left(\sum_{i=1}^{n}\frac{1}{p_i}\right)}\left[1 - 2p_i\sum_{i=1}^{n}\frac{1}{p_i}\right] - 2 < 0 \text{ since } \left(1 - 2p_i\sum_{i=1}^{n}\frac{1}{p_i}\right)$$

Thus, $\psi''(P)_n < 0$, which shows that $\psi_n(P)$ is a cor

Thus, we have the following properties:

(i) $\psi_n(P)$ is permutationaly symmetric as it does not changes if $p_1, p_2, p_3, \ldots, p_n$ are rearranged among themselves. This property is desirable since the labeling of the outcomes should not affect the entropy.

(ii) Since $\frac{1}{p_i}$ is continuous function of $p_1, p_2, p_3, \ldots, p_n$ for $0 < p_i < 1$, $\psi_n(P)$ is also continuous everywhere

(iv) $\xi_n(P)$ is a concave function of $p_1, p_2, p_3, \ldots, p_n$. This is very useful property since a local maximum will also be the global maximum for a concave function.

(V) For maximum value, we consider the following Lagrangian:

$$L = -\log\left[\sum_{i=1}^{n}\frac{1}{p_i}\right] - \lambda\left[\sum_{i=1}^{n} p_i - 1\right]$$

Thus ,we have $\frac{\partial L}{\partial p_1} = \frac{1}{\sum_{i=1}^{n}\frac{1}{p_i}}\ \frac{1}{p_1^2} - 2p_i - \lambda$, $\frac{\partial L}{\partial p_2} = \frac{1}{\sum_{i=1}^{n}\frac{1}{p_i}}\ \frac{1}{p_2^2} - 2p^2 - \lambda$, …, $\frac{\partial L}{\partial p_i} = \frac{1}{\sum_{i=1}^{n}\frac{1}{p_i}}\ \frac{1}{p_i^2} = -2p_i - \lambda$

Now, $\frac{\partial L}{\partial p_i} = 0$ gives that $\frac{\partial L}{\partial p_1} = \frac{\partial L}{\partial p_2} \ldots .= \frac{\partial L}{\partial p_i}$

This is possible if and only if

Also $\sum_{i=1}^{n} p_i = 1 \Rightarrow n.p_i = 1 \Rightarrow p_i = \frac{1}{n}$

Thus, the maximum value of $\psi_n(P)$ will exist at $p_i = \frac{1}{n}$ and is given by

$$\left[\psi_n(P)\right]_{max} = -2\log n - \frac{1}{n}$$

Hence, we see that $\psi_n(P)$ introduced in equation (2.10) satisfies all the essential properties of an information measure, it is a new measure of information. It is thus concluded that for the given values of arithmetic mean, harmonic mean and the standard deviation, the information content of the probability distribution can be obtained. Next, with the help of the data, we have presented the measure (2.10) graphically. For this purpose, we have fixed n =2, then for different probabilities, we have computed different values of $\psi_n(P)$ as shown in table- 2.1.

Table-2.1

p_1	p_2	$\xi_n(P)$
0.05	0.95	-5.3009
0.15	0.85	-3.7164
0.30	0.70	-2.8315
0.40	0.60	-2.5789
0.50	0.50	-2.5000
0.60	0.40	-2.5789
0.70	0.30	-2.8315
0.80	0.20	-3.3239
0.90	0.10	-4.2939
0.95	0.05	-5.3009

Next, we have presented the values of $\psi_n(P)$ graphically and obtained the following Fig.-2.2 which shows that the measure introduced in equation (2.10) is a concave function.

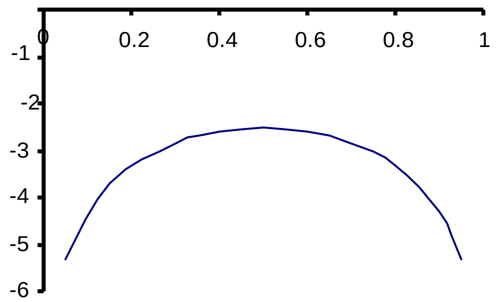

Fig.-2.1

(I) Information Measure in terms of Arithmetic Mean, Harmonic Mean, Geometric Mean and Standard Deviation

From equations (2.1) and (2.2), we have

$$\frac{G}{M} = n\,(p_1\, p_2\, p_3 \dots p)^{1}$$

OR

$$\frac{1}{2}\sum_{i=1}^{n} \log p_i = \frac{1}{2^n}\, n\log\left(\frac{G}{nM}\right) \qquad (2.11)$$

Adding equations (2.6), (2.8) and (2.11), we get

$$\frac{\sigma^2 + M^2}{2^n n M^2} + \frac{H}{M} + \frac{1}{2^n}\, n\log + \left(\frac{G}{nM}\right) = \frac{1}{2^n}\sum_{i=1}^{n} p_i^2 + \frac{n^2}{\sum_{i=1}^{n}\frac{1}{p_i}} + \frac{1}{2^n}\sum_{i=1}^{n} \log p_i \qquad (2.12)$$

Now, we introduce an information theoretic measure depending upon arithmetic mean M, harmonic mean H, geometric mean G and variance σ^2. This measure is given by

$$\xi(P) = \frac{1}{2^n}\sum_{i=1}^{n} p_i^2 + \frac{n^2}{\sum_{i=1}^{n}\frac{1}{p_i}} + \frac{1}{2^n}\sum_{i=1}^{n} \log p_i; \qquad n \geq 2, \quad 0 < p_i < 1 \qquad (2.13)$$

We shall prove that the R.H.S. of equation (2.13) is an information measure. To prove this, we study its following properties:

(i) $\xi_n(P)$ is continuous function of $p^i \ \forall 0 < p^i < 1$

(ii) $\xi_n(P)$ is permutationally symmetric function of $p^i \ \forall 0 < p^i < 1$

(iii) $\xi_n(P)$ may be positive or negative. Usually an information measure is supposed to be non- negative but Burg's [1] measure of entropy gives negative value. The measure (2.13) includes Burg's [1] measure of entropy, thus it can take both values.

(iv) **Maximum Value:** To obtain, the maximum value of the entropy measure (2.13), we consider the following Lagrange's function:

$$L= \frac{1}{2^n}\sum_{i=1}^{n} p_i^2 + \frac{n^2}{n \quad 1} + \frac{1}{2^n}\sum_{i=1}^{n} \log p_i - \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$$

Thus ,we have $\frac{\partial L}{\partial p_1} = \frac{2p_1}{2^n} + \frac{n^2}{(p_1)^2}\left(\sum_{i=1}^{n}\frac{1}{p_i}\right)^2 - \frac{1}{2^n p_1} - \lambda$

Similarly,

$$\frac{\partial L}{\partial p_2} = \frac{2p_2}{2^n} + \frac{n^2}{(p_2)^2}\left(\sum_{i=1}^{n}\frac{1}{p_i}\right)^{-2} - \frac{1}{2^n p_2} \quad \lambda, \ldots, \quad \frac{\partial L}{\partial p_n} = \frac{2p_2}{2^n} + \frac{n^2}{(p_n)^2}\left(\sum_{i=1}^{n}\frac{1}{p_i}\right)^{-2} - \frac{1}{2^n p_n} - \lambda$$

For maximum value, we put

$$\frac{\partial L}{\partial p_1} = \frac{\partial L}{\partial p_2} = \frac{\partial L}{\partial p_3} = \frac{\partial L}{\partial p_n} = \text{which gives } p_1 = p_2 = \ldots = p_n$$

Also using $\sum_{i=1}^{n} p_i = 1$, we get $p_i = \frac{1}{n}$ Thus, the maximum value arises when the distribution is uniform. Further, the maximum value is

$$f(n) = \frac{1}{n2^n} + n^2 - \frac{1}{2^n} n \log n$$

Also

$$f'(n) = 2n - \frac{1}{2^n}\left(\frac{1+n}{n^2} + \log n(n-1) - 1\right) > 0 \qquad \forall n \geq 2$$

Hence, the maximum value is an increasing function of n.

(v) Concavity: To study its concavity, we have

$$\xi^{n}_{n} = -2\left[\frac{n^2}{P_i^4\left(\sum_{i=1}^{n}\frac{1}{p_i}\right)^3}\left\{p_i\sum_{i=1}^{n}\frac{1}{p_i} - 1\right\} - \frac{1}{2^n}\right] - \frac{1}{2^n p_i^2} < 0$$

which shows that $\xi_n(P)$ is concave. Thus, we see that the measure introduced in equation (2.13) satisfies all the essential properties of being an information measure. Hence, we conclude that $\xi_n(P)$ is another new measure of information. It is thus concluded that for the given values of arithmetic mean, harmonic mean, geometric mean and the standard deviation, the information content of the probability distribution can be obtained. Next, with the help of the data, we have presented the measure (2.13) graphically and obtained the following Fig.-2.2 which shows that the measure introduced in equation (2.13) is a concave function.

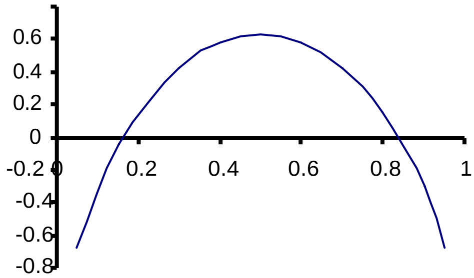

Fig.-2.2

References

1. Burg, J.P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. In: Childrers, D.G.(ed). Modern Spectral Analysis. pp130-131.
2. Havrada, J.H. and Charvat, F. (1967). Quantification methods of classification process: Concept of structural □-entropy. Kybernetika **3**: 30-35.
3. Herremoes, P.(2006). Interpretations of Renyi entropies and divergences. Journal of Physics A **365**: 57-62.
4. Kapur, J.N. (1967). Generalized entropy of order □ and type □. Mathematics Seminar **4**: 79-84.
5. Renyi, A. (1961). On measures of entropy and information. Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability **1**: 547-561.
6. Shannon, C. E.(1948). A mathematical theory of communication. Bell System Technical Journal **27**: 379- 423, 623-659.
7. Sharma, B.D. and Taneja, I.J. (1975). Entropies of type (□,□) and other generalized measures of information theory. Metrica **22**: 202-215.
8. Simpson, E.H. (1949). Measurement of diversity. Nature **163**: 688.
9. Wiener, N. (1949). The extrapolation, interpolation and smoothing of stationary time series with engineering applications. John Wiley and Sons, New York.
10. Zadeh, L.A. (2005). Towards a generalized theory of uncertainty (GTU)- an outline. Information Sciences **172**: 1-40.