International Journal of Mathematical Archive-6(6), 2015, 202-208
Available online through www.ijma.info ISSN 2229 – 5046

A REVIEW OF SOME APPLICATIONS OF CLASSICAL PROBABILISTIC INFORMATION MEASURES

SURENDER SINGH*

School of Mathematics, Faculty of Sciences,
Shri Mata Vaishno Devi University, Sub Post Office, Katra-182320 (J & K), India.

(Received On: 06-06-15; Revised & Accepted On: 30-06-15)

ABSTRACT

Information theory, which was developed as a branch of communication engineering mainly by Hartley (1928) and Shannon (1948), has its roots in statistical thermodynamics. It has been applied in many fields outside communication engineering. In the present communications some application of classical Shannon measure and generalized entropy measures have been reviewed.

Key Words: *Information Measure, Maximum Entropy Principle, Generalized Information Measures.*

2010 Mathematics Subject Classification: *94A17.*

1. INTRODUCTION

Consider a set E of mutually exclusive events E_i (i = 1... n) each of which has the probability of occurrence p_i, so that the p_i s add up to unity. The *information content* of the occurrence of event E_i is defined [28]:

$$Inf\,(E_i) = -\log p_i = \log \frac{1}{p_i}$$

The expected information content of an event from our set of n events, the entropy of the set E, is defined [28]:

$$H(P) = \sum_{i=1}^{n} p_i \; Inf\,(E_i) = \sum_{i=1}^{n} p_i \; \log \; p_i \tag{1.1}$$

H(P) is always non negative. Its maximum value depends on n. It is equal to $\log n$ when all p_i are equal. A general *communication system* has a *source* which generates *messages* symbol by symbol and chooses symbols with given probabilities. The symbols are transmitted through a *channel* in which *noise* may disturb them. The messages are received at *destination*.

Let $p(x_i) =$ the probability that the source will generate and send symbol x_i; $p(m_j) =$ the probability that the destination will receive symbol m_j; $p(x_i, m_j) =$ the probability that symbol x_i will be sent and symbol m_j will be received; $p(m_j \mid x_i) =$ the probability that symbol m_j will be received when symbol x_i has been sent; $p(x_i \mid m_j) =$ the probability that symbol x_i was sent if symbol m_j has been received.

Now using (1.1) we can define the entropies of the communication system:

$$H(x) = \sum_i p(x_i) \log \frac{1}{p(x_i)} \; ;$$

$$H(m) = \sum_i p(m_j) \log \frac{1}{p(m_j)};$$

$$H(x,m) = \sum_i \sum_j p(x_i, m_j) \log \frac{1}{p(x_i, m_j)};$$

Corresponding Author: Surender Singh*

$$H(m\,|\,x)=\sum_i\sum_j p(x_i)p(m_j\,|\,x_i)\log\frac{1}{p(m_j\,|\,x_i)};$$
$$H(x\,|\,m)=\sum_i\sum_j p(m_j)p(x_i\,|\,m_j)\log\frac{1}{p(x_i\,|\,m_j)}.$$

System in which

$$p(m_j\,|\,x_i)=\begin{cases}1 & when \quad i=j\\ 0 & when \quad i\neq j\end{cases}$$

is called *noiseless systems.* Consequently, in noiseless systems $H(m\,|\,x)=H(x\,|\,m)=0$ and also $H(x)=H(m)$. In noisy communication systems the conditional entropy of the symbols sent when the received symbols are known $H(x\,|\,m)$, has been taken to measure the *missing information* due to noise in the channel. $H(x\,|\,m)$ is called the *equivocation of the channel.* The rate of actual transmission of the information can be defined in three equivalent ways:

$$\begin{aligned}R&=H(x)-H(x\,|\,m)=H(m)-H(m\,|\,x)\\&=H(x)+H(m)-H(x,m)\end{aligned}$$

R is always non negative. Its minimum is reached when $p(x_i,m_j)=p(x_i)p(m_j)$, i.e when the symbols sent and received are independent. This means that there is maximum noise in the system. The maximum of R is $H(x)$ and is reached when the system is noiseless.

Mutual Information

Mutual Information between two random variables X and Y can be defined in following three equivalent ways

$$\begin{aligned}I(X,Y)&=H(X)-H(X\,|\,Y)\\&=H(Y)-H(Y\,|\,X)\\&=H(X)+H(Y)-H(X,Y)\end{aligned}\qquad (1.2)$$

$I(X,Y)$ can be interpreted as a measure of the dependence between Y and X. When X and Y are independent, $I(X,Y)$ is minimum, that is, $I(X,Y)$ =0.

Based upon this property, Mutual Information (MI) was introduced as a similarity measure in various applications.

In 1961, Renyi [26] pointed out in his fundamental paper on generalized information measure that in other sorts of problems other quantities may serve just as well or even better than measures of information. The basic idea behind Renyi's generalization is that an expression for an entropy should be a mean, and therefore, he uses a well known mathematical idea that the linear mean, though most widely used, is not the only possible way of averaging, however, one can define the mean with respect to an arbitrary function. Here one should be aware that, to define a 'meaningful' generalized mean, one has to restrict the choice of functions to continuous and monotone functions. Following the above idea, once we replace the linear mean with generalized means, we have a set of information measures corresponding to a continuous and monotone function. Renyi 's[26] celebrated measure is

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

Shannon,s measure is a limiting case of Renyi,s measure. After Renyi, voluminous literature on generalization (one, two, three and four parametric) of information measures developed.

But utility of these additional parameter(s) remained missing in most of the work on generalization. In this review, some discussion on pragmatic aspect of parameter(s) is also given.

Since in most of interdisciplinary applications maximum entropy principle plays an important role. Therefore, in section 2, principle of maximum entropy has been recalled. Section 3 reviews the applications of non-parametric information measure (entropy measure) and in section 4 some insight on application of generalized (parametric) information measure is given.

2. MAXIMUM ENTROPY PRINCIPLE

We know the finite discrete case of Shannon's entropy, which is given by Boltzmann's H-function taking from statistical mechanics. We define the entropy function as in (2.1) .Where $H(P)$ is the amount of uncertainty contained by the probability distribution P, is known as system's entropy function. The property of Shannon's entropy is that

$$H(P)=\sum_{i=1}^{n} p_i \log p_i = H(p_1,p_2,...p_n)\le H\left(\frac{1}{n},\frac{1}{n},...\frac{1}{n}\right) \tag{1.1}$$

By this expression we notice that the uniform distribution is the most uncertain one when no constraint is imposed on the probability distribution. This relation is equal to Laplace's principle of insufficient reason, which implies that the most reasonable strategy consists in attaching the same probability to different outcomes when we have no additional information about them. Jaynes [12] extended Laplace's principle of sufficient reason by introducing the PME, which proves to be a non-linear entropy function (1.1) to be maximizes subject to the constraints

$$E(f)=\sum_{k=1}^{n} f_k p_k \tag{2.2}$$

Where f_k , (k = 1,2,...,n) are weights which give information about moments in certain cases.

Also,

$$\sum_{k=1}^{n} p_k = 1 \tag{2.3}$$

We select the probability distribution P = (p_1, p_2,..., p_n) that maximizes the corresponding entropy relation (1.1) subject to the given constraints.

3. APPLICATIONS OF PROBABILISTIC INFORMATION MEASURES

Beyond the theory of communications, applications of probabilistic information measures could be seen in finance, marketing, operations research, actuarial sciences, image processing, sociology, biology, ecology, population studies and many more. In this section some of the applications have been recalled.

3.1 Application in finance and trade

Problem: Could a mathematical formula be derived in betting on horse races?

J. Kelly in his article "A new Interpretation of Information rate" [16] provided a solution know as Kelly Formula or Kelly Criterion became the basis of many betting systems and also paved the way for various strategies in financial options trading. According to Kelly random variable X denotes the outcome of each horse race is represented by a random variable *X* and gambler has some side information *Y* relevant to the outcome of the race, then under some condition of odds, the mutual information *I*(*X; Y*) is the difference in growth rates of the wealth with and without side information. The side information can be considered to be a set of symbols received over a communication channel that the gambler uses to make bets on the transmitted symbols. The growth rate is defined as

$$G=\lim_{n\to\infty}.\frac{1}{N}\log\frac{V_N}{V_0} \tag{3.1}$$

where V_N is the gamblers capital after N bets and V_0 is the starting capital. An optimal gambler always tries to maximize the growth rate (G_{max}).

Problem: Could the expected portfolio return be maximized from an investment in share market?

Wilson [34] proposed mean entropy model using concepts of individual entropy, joint entropy and conditional entropy.

In international trade, distribution pattern of industrial agglomeration can be studied in information theoretic framework.

Let k_{ij} stand for the trade flow from region i to region j $(i, j = 1,2,...,n)$. Then

$X_i=\sum_j k_{ij}=$ Total export of the region *i;* $M_J=\sum_i k_{ij}=$ Total import of region

$X=\sum_i X_i=\sum_j M_j=$ Total world trade.

While analyzing the international trade, we calculate several statistics to describe them. As for example we analyze the changes *Export shares*, which we present in the form of the export vector $\{X_i / X\}$; *Import shares*, which we present in the form of the import vector $\{M_j / X\}$; *Trade patterns*, which we present in the form of the trade-flow matrix $\{k_{ij} / X\}$; *Export patterns*, which we present in the form of the conditional export vector $\{k_{ij} / X_i\}$; *Import patterns*, which we present in the form of the conditional import vector $\{k_{ij} / M_j\}$;. Entropy analysis of international trade using empirical data is done by Pulliainen [25]. H(m | x) and H(x | m) are weighted averages of the conditional entropies calculated for individual regions, and they measure the average uncertainty of destination and origin, given the region of origin and destination, respectively. These entropies also measure the evenness or concentration of the distributions of the elements in the trade-flow matrix and vectors. The larger the entropies the more even are the distributions in question; the smaller the entropies the more concentrated they are. Jacquemin and Charles [11] studied the relation between diversification and corporate growth using measure of entropy.

3.2 Application in market research

Problem: The most fundamental question that can be asked about consumer choice behavior, is that whether, there exist causes and explanations for all behaviors in some fundamental sense?

Since individual consumer choice behavior is characterized by some randomness, the underlying fact which guides most of the research in individual consumer behavior is that, in principle, behavior is caused and can therefore be explained.

Keeping in view the above fact, Herniter [10] maximized the shannon's measure of entropy to estimate the probability of customer changing from from one brand of a product to another brand, Bass [2] criticized him for inflexibility of his results, since Hertiner [10] probability of transition was same whether products was soap or alcohal, or whether community belongs to Africa or Asia or U.S.A. Later, Kapur *et al.* [15] remedied this defect by using Renyi's entropy measure, which depends on one parameter.

3.3 Application in operations research

In operations research there are two fundamental problems one is optimization and second system modelling. Kapur [14] found different objective functions for different entropies. Presence of one, two or three parameters provides flexibility for choice of best entropy measure for the optimization problem. It is very important to note that the principle of maximum entropy is at the core of every optimization problem.

In system modelling the concept of entropy seems to have four perspectives, the first three essentially subjective and the last objective:

(i) Entropy can be considered as probability (to deal with the notion of uncertainty).
(ii) Entropy can be considered as a statistic of a probability distribution.
(iii) Entropy can be considered as the negative of a Bayesian log-likelihood function.
(iv) Entropy as a measurable system property (to deal with randomness and disorder).

Entropy has three types of utility in model building:

(i) For hypothesis generation and as a means of achieving consistency within models of certain types of rather complex situations,
(ii) For interpreting models and theories, and
(iii) For the development of dynamic system models (as distinctive from equilibrium models to be used in a comparative-static way).

Wilson [34] gave a nice framework for building models of urban and regional systems used for planning purposes using the concept of entropy.

In optimization problems, queuing theory is also a versatile discipline. Queues are formed whenever individuals have to wait for receiving service. A queue may be formed either by human beings or by machines. Waiting in queues may be very costly in time and money and a scientific study of queues is essential in order to reduce cost for given waiting time or reduce waiting time for given cost. Queuing theory is concerned with the following problems:

Situation: The situation is described by probability distributions of the inter-arrival time and the service time, the queuing discipline, and other features like limited waiting space, persons not joining a queue if they find it too long and persons losing patience and leaving the queue after waiting in it for some time.

Problem: The problem is to find performance distributions like, queue size distribution, waiting time distribution and busy period distribution.

The maximum entropy principle is applicable to select appropriate probability distributions for a queuing situation. For example, some moments of the inter-arrival time or service time or queue size waiting time distribution may be given. In each of these cases, we can use the maximum entropy principle to select a corresponding probability distribution based only on the available information. Sometimes the distributions obtained by the usual methods applied in queuing theory are rather complicated. In such cases the principle of maximum entropy can be used to approximate these distributions by simpler ones of the exponential family. The contributions of Guiasu [7] and Kouvatsos [17] contain examples showing the role of maximum entropy in queuing theory. Artalezo and Lopez-herrero [1] applies information theoretic methodology for systems modeling to investigate the probability density function of busy period of M/G/1 vacation model operating under the N-, T- and D- policies. Parkash and Gandhi [23] study variations of different entropy measures in steady and non-steady processes of queuing theory.

3.4 Application in population studies

The entropy of population measures the variability of the contribution of the different age classes to the stationary age distribution. Demetrius [6] studied the problem of natural selection using entropy measure and obtained a result analogous to Fisher fundamental theorem of natural selection. Maasumi [18] studied a class of evaluative statistics for the measurement of multidimensional inequality and their corresponding Social Welfare Functions.

3.5 Application to measure the diversity

Diversity is a measure of the compositional complexity of an assemblage. It plays a central role in community ecology and conservation biology. If n is the number of species in the assemblage and ith species have relative abundance p_i then the shannon's entropy $H(P) = -\sum_{i=1}^{n} p_i \log p_i$ gives the uncertainty in the species identity of a randomly chosen individual in the assemblage. Burbea and Rao[3] measured the total diversity in a mixed distribution in terms of average diversity within the distribution.

4. APPLICATIONS OF GENERALIZED INFORMATION MEASURES

Owing to various theoretical and practical reasons, related to various branches of study, it appears that Shannon entropy is the most important and useful measure of uncertainty, and, consequently a new entropy will be of interest if and only if it has some advantage with respect to the former. This section deals some applications where the generalized entropies have some advantage.

4.1 Application in Non extensive Information Systems

Tsallis[32] proposed nonextensive formalism of concept of entropy by proposing a generalized q-entropy S_q which satisfies

$$S_q(A+B) = S_q(A) + S_q(B) + \tau(q)\, S_q(A) S_q(B), \tag{4.1}$$

Where $\tau(q)$ is called *entropic index*. For Tsallis entropy $\tau(q) = 1-q,$

we have

$$S_q(A+B) = S_q(A) + S_q(B) + (1-q) S_q(A) S_q(B), \tag{4.2}$$

(4.2) is called pseudo-additivity property.

When $q \to 1$, (4.2) reduces to

$$S(A+B) = S(A) + S(B), \tag{4.3}$$

Which is Shannon's additivity. The deviation of '*q*' from unity measures the degree of non extensivity of the entropies. Thus, generalized entropies are useful for dealing the *nonextensive systems* (systems where the output is not equal to the sum of inputs supplied). Generalized entropies are also useful in Statistical physics for the purpose of analyzing multifractal systems.

4.2 Application in Chaos Theory

Benoit B.Mandelbrot [19] determined "fractal": A rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced/size copy of the whole. And fractal dimension is the dimension number of fractal, and it metrics the ability of system filled space (compact) or crannies (osteoporosis), and characterizes the system disorder.

Generally, fractal dimension is calculated by:

$$D_f = -\frac{\ln N(\varepsilon)}{\ln(\varepsilon)} \tag{4.4}$$

ε is measuring scale and $N(\varepsilon)$ is number of parts under ε.Generalized dimension was proposed by Shixian[29] as follows:

$$D_\alpha = -\lim_{\varepsilon \to 0} . \frac{H_\alpha(\varepsilon)}{\ln(\varepsilon)} \tag{4.5}$$

Where $H_\alpha(\varepsilon)$ is Renyi's entropy.

4.3 Application to Provide Flexibility and Subjectivity in Information System

Jumarie [13] have shown that Renyi's entropy $H^R(X)$ can be utilized to model the effect of the subjectivity in the information systems.

4.4 Application in Statistical Mechanics

Well known probability distributions of statistical mechanics like Boltzaman-Gibbs (BG), Bose-Einestien (BE), Fermi-Dirac (FD) distributions are obtain by maximizing the generalized entropies. More detail are given in Jaynes [12], Kapur[14], Tsallis and Brigatti [31] and Niven[22].

4.5 Application of Parameter for Scaling the Information Measure

The generalized entropies, with the Shannon entropy as a special case, are almost consistent for different values of α. Increasing α makes the measure's values span a smaller interval. This means that as α increases, the measures become coarse and their discriminating power decreases. So, the parameter α is used for scaling the measure. Some applications of this type are Neemuchwala *et al.* [21] and Mohanalin *et al.* [20]. Generalized information measures proposed in literature contains one, two or more parameters. These parameters give certain flexibility in applications and their values have ultimately to be determined from the data itself.

4.5 Application in Image Processing

In image processing, two types of problem exists which have been studied in information theoretic environment.

Problem: To Align images from different modalities (Image registration). This is very useful in the diagnoses of diseases.

Intensity based image registration is one of the most popularly used methods for automatic multimodality image registration. Recently, various improvements have been suggested, ranging from variation in the similarity metrics to improvement in the interpolation techniques. As a similarity measure, mutual information was pioneered both by Collignon et al. [4], and by Viola and Wells [33]. Applied to rigid registration of multimodality images, mutual information showed great promise and within a few years it became the most investigated measure for medical image registration.

Problem: To Divide an image into meaningful structures (image segmentation). This is often an essential step in image analysis, object representation, visualization and many other image processing tasks.

In recent years, multi scale techniques have gained a lot of attention in the image processing community. Typical examples are pyramid and wavelet decompositions. They represent images at a small number of scales and have proven their use for many image processing tasks. It is not difficult to see that the generalized entropies, the multi fractal spectrum, the gray-value moments and the gray-value histogram itself are equivalent representations: they can be transformed into each other by one-to-one mappings. The relation between gray-value moments and the gray-value histogram is given in [30] The equivalence of the multifractal spectrum and the generalized entropies is discussed in [9]. Further [30] have illustrated that the generalized entropies can be used to perform size measurements for periodic textures. This is not possible with the Shannon–Wiener entropy. Further applications can be seen in [24], [27] and [5].

REFERENCES

1. Artalezo, J. R., and Lopez-Herrero, M.J., Entropy Maximization and Busy Period of Some Single-Server Vacation Models, Rairo Operations Research, 38(2004), 195-213.
2. Bass, F.M., The Theory of Stochastic Preference and Brand Switching, J. Market Res., 11(1974), 1-20.

3. Burbea, J., and Rao, C.R., On the convexity of some divergence measures based on entropy functions,IEEE Trans. Inf. Theory,IT-28(3)(1982),489–495.
4. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G., Automated multimodality medical image registration using information theory, Proc. int. conf. information processing medical imaging: Computational imaging and vision,(1995), 263-274.
5. De Albuquerque, M.P., Esquef, I. A. and Gesualdi Mello, A. R., Image thresholding using Tsallis entropy, Pattern Recognition Letters, 25(2004),1059–1065.
6. Demetrius, L., Demographic parameters and natural selection, Proc. Nat. Acad. Sci., USA, Vol. 71(12) (1974),4645-4647.
7. Guiasu, S., Maximum Entropy Condition in Queuing Theory, The Journal of the Operational Research Society, 37(3)(1986), 293-301.
8. Hartley, R. V. L., Transmission of Information, Bell System Technical Journal, 7(1928), 535-563.
9. Hasley, T. C., Jensen, M. H., Kadanoff, L.P., Procaccia, I. and Shraiman, B. I., Fractal measures and their singularities: The characterization of strange sets, Phys. Rev. A, 33(1986), 1141–1151.
10. Herniter, J. D., Model of Brand Purchase Behavior, J. Market Res., 10(1973), 361-375.
11. Jacquemin, A. P. and Berry, C. H., Entropy measure of diversification and corporate growth, The Journal of Industrial Economics, 27(4) (1979), 359-369.
12. Jaynes, E.T., Information theory and statistical mechanics, Phys. Rev., 106(1957), 620-630.
13. Jumarie, G., Relative Information: Theories and applications, Springer-Verlag, New York, 1990.
14. Kapur, J. N., Measures of uncertainty mathematical programming and physics, Department of Mathematical Sciences, Mellon College of Science, Paper 162(1970).
15. Kapur, J.N., Bector, C.R. and Kumar, U., On the generalized entropy model for brand switching, Naval Res. Log. Quart., 31(1984), 183-198.
16. Kelly, J., A new interpretation of information rate, Bell Syst. Tech. J, July 1956.
17. Kouvatsos,D. D., A Maximum Entropy Analysis of the G/G/1 Queue at Equilibrium, The Journal of the Operational Research Society,39(2)(1988),183-200.
18. Maasoumi, E., The measurement and decomposition of multidimensional inequality, Econimetrica, 54(4) (1986), 991-997.
19. Mandelbrot, B.B., Intermittent turbulence in self- similar cascades: Divergence of higher moments, J. Fluid Mech., 62(2) (1974), 331-358.
20. Mohanalin, B., Kalra, P. K. and Kumar, N., An automatic image registration scheme using Tsallis entropy, Biomedical Signal Processing and Control, 5(2010), 328–335.
21. Neemuchwala, H., Heroa, A. and Carsona, P., Image matching using alpha-entropy measures and entropic graphs, Signal Processing, 85(2005), 277–296.
22. Niven, R. K., Exact Maxwell–Boltzmann, Bose–Einstein and Fermi–Dirac statistics, Physics Letters A, 342(2005), 286–293.
23. Parkash, O. and Gandhi, C. P., Variation of uncertainty in steady and Non-steady processes of queuing Theory, World Academy of Science, Engineering and Technology, 51(2011),1396-99.
24. Pavesic, N. and Ribaric, S., Gray level thresholding using the Havrda and Charvat entropy, 10th Mediterranean Electrotecnical Conference, MEleCon, 2(2000), 631-634.
25. Pulliainen, K., Entropy-Measures for International Trade, The Swedish Journal of Economics, 72(1), Problems in International Trade (Jan., 1970), 40-53.
26. Renyi, A., On measures of entropy and information, Proc. 4th Berkeley Symp. Math. Stat. Probab. 1, (1961), 547-561.
27. Sahoo, P.K. and Arora, G., A thresholding method based on two-dimensional Renyi's entropy, Pattern Recognition, 37(2004), 1149 – 1161.
28. Shannon,C. E., A Mathematical Theory of Communication, Bell System Technical Journal, 27, 1948, 379-423, 623-656.
29. Shixian, Q., The relation between fractal dimension and entropy, Chinese Journal of High Pressure Physics, 7(2)(1993),127-132(in Chinese).
30. Sporring, J. and Weickert, J., Information measures in scale-spaces, IEEE Transactions On Information Theory, 45(3) (1999), 1051-1058.
31. Tsallis, C. and Brigatti, E., Nonextensive statistical mechanics: A brief introduction, Available at: arXiv:cond-mat/0305606v1 [cond-mat.stat-mech] 26 May 2003.
32. Tsallis,C., Possible generalization of Boltzmann-Gibbs statistics, J. Statist. Phys., 52 (1988), 479–487.
33. Viola, P. and Wells, W. M., Alignment by maximization of mutual information, Proc. Int. Conf. on Computer Vision, Cambridge, CA, 1995, 16-23.
34. Wilson, A. G., The use of concept of entropy in system modelling, Operations Research Quarterly, 21(2) (1970), 247-265.

Source of support: Nil, Conflict of interest: None Declared