SHANNON ENTROPY: AXIOMATIC CHARACTERIZATION AND APPLICATION

C. G. CHAKRABARTI AND INDRANIL CHAKRABARTY

We have presented a new axiomatic derivation of Shannon entropy for a discrete probability distribution on the basis of the postulates of additivity and concavity of the entropy function. We have then modified Shannon entropy to take account of observational uncertainty. The modified entropy reduces, in the limiting case, to the form of Shannon differential entropy. As an application, we have derived the expression for classical entropy of statistical mechanics from the quantized form of the entropy.

1. Introduction

Shannon entropy is the key concept of information theory [12]. It has found wide applications in different fields of science and technology [3, 4, 5, 7]. It is a characteristic of probability distribution providing a measure of uncertainty associated with the probability distribution. There are different approaches to the derivation of Shannon entropy based on different postulates or axioms [1, 8].

The object of present paper is to stress the importance of the properties of additivity and concavity in the determination of functional form of Shannon entropy and its generalization. The main content of the paper is divided into three sections. In Section 2, we have provided an axiomatic derivation of Shannon entropy on the basis of the properties of additivity and concavity of entropy function. In Section 3, we have generalized Shannon entropy and introduced the notion of total entropy to take account of observational uncertainty. The entropy of continuous distribution, called the differential entropy, has been obtained as a limiting value. In Section 4, the differential entropy along with the quantum uncertainty relation has been used to derive the expression of classical entropy in statistical mechanics.

2. Shannon entropy: axiomatic characterization

Let Δ_n be the set of all finite discrete probability distribution

$$P = \left\{ (p_1, p_2, \ldots, p_n),\ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1 \right\}. \tag{2.1}$$

In other words, P may be considered as a random experiment having n possible outcomes with probabilities $(p_1, p_2,\ldots, p_n)$. There is uncertainty associated with the probability distribution P and there are different measures of uncertainty depending on different postulates or conditions. In general, the uncertainty associated with the random experiment P is a mapping [9]

$$H(P): \Delta_n \longrightarrow \mathbb{R}, \tag{2.2}$$

where $\mathbb{R}$ is the set of real numbers. It can be shown that (2.2) is a reasonable measure of uncertainty if and only if it is a Shur concave on Δ_n [9]. A general class of uncertainty measures is given by

$$H(p) = \sum_{i=1}^{n} \phi(p_i), \tag{2.3}$$

where $\phi : [0,1] \to \mathbb{R}$ is a concave function. By taking different concave function defined on $[0,1]$, we get different measures of uncertainty or entropy. For example, if we take $\phi(p_i) = -p_i \log p_i$, we get Shannon entropy [12]

$$H(P) = H(p_1, p_2,\ldots, p_n) = -k \sum_{i=1}^{n} p_i \log p_i, \tag{2.4}$$

where $0 \log 0 = 0$ by convention and k is a constant depending on the unit of measurement of entropy. There are different axiomatic characterizations of Shannon entropy based on different set of axioms [1, 8]. In the following, we will present a different approach depending on the concavity character of entropy function. We set the following axiom to be satisfied by the entropy function $H(P) = H(p_1, p_2,\ldots, p_n)$.

AXIOM 1. *We assume that the entropy $H(P)$ is nonnegative, that is, for all $P = (p_1, p_2, \ldots, p_n)$, $H(P) \geq 0$. This is essential for a measure.*

AXIOM 2. *We assume that generalized form of entropy function (2.3) is*

$$H(P) = \sum_{i=1}^{n} \phi(p_i). \tag{2.5}$$

AXIOM 3. *We assume that the function ϕ is a continuous concave function of its arguments.*

AXIOM 4. *We assume the additivity of entropy, that is, for any two statistically independent experiments $P = (p_1, p_2,\ldots, p_n)$ and $Q = (q_1, q_2,\ldots, q_m)$,*

$$H(PQ) = \sum_{j} \sum_{\alpha} \phi(p_j q_\alpha) = \sum_{j} \phi(p_j) + \sum_{\alpha} \phi(q_\alpha). \tag{2.6}$$

Then we have the following theorem.

THEOREM 2.1. *If the entropy function $H(P)$ satisfies Axioms 1 to 4, then $H(P)$ is given by*

$$H(P) = -k \sum_{i=1}^{n} p_i \log p_i, \tag{2.7}$$

where k is a positive constant depending on the unit of measurement of entropy.

Proof. For two statistically independent experiments, the joint probability distribution $p_{j\alpha}$ is the direct product of the individual probability distributions

$$p_{j\alpha} = p_j \cdot q_\alpha. \tag{2.8}$$

Then according to the axiom of additivity of entropy (2.6), we have

$$\sum_j \sum_\alpha \phi(p_j \cdot q_\alpha) = \sum_j \phi(p_j) + \sum_\alpha \phi(q_\alpha). \tag{2.9}$$

Let us now make small changes of the probabilities p_k and p_j of the probability distribution $P = (p_1, p_2, \ldots, p_j, \ldots, p_k, \ldots, p_n)$ leaving others undisturbed and keeping the normalization condition fixed. By the axiom of continuity of ϕ, the relation (2.9) can be reduced to the form

$$\sum_\alpha q_\alpha [\phi'(p_j \cdot q_\alpha) - \phi'(p_k \cdot q_\alpha)] = \{\phi'(p_j) - \phi'(p_k)\}. \tag{2.10}$$

The right-hand side of (2.10) is independent of q_α and the relation (2.10) is satisfied independently of p's if

$$\phi'(q_\alpha \cdot p_j) - \phi'(q_\alpha p_k) = \phi'(p_j) - \phi'(p_k). \tag{2.11}$$

The above leads to the Cauchy functional equation

$$\phi'(q_\alpha \cdot p_j) = \phi'(q_\alpha) + \phi'(p_j). \tag{2.12}$$

The solution of the functional equation (2.12) is given by

$$\phi'(p_j) = A \log p_j + B \tag{2.13}$$

or

$$\phi(p_j) = A p_j \log p_j + (B - A) p_j + C, \tag{2.14}$$

where A, B, and C are all constants. The condition of concavity (Axiom 3) requires $A < 0$ and let us take $A = -k$ where $k(>0)$ is positive constant by Axiom 1. The generalized entropy (2.5) then reduces to the form

$$H(P) = -k \sum_j p_j \log p_j + (B - A) + C \tag{2.15}$$

or

$$H(P) = -k \sum_j p_j \log p_j, \tag{2.16}$$

where constants $(B - A)$ and C have been omitted without changing the character of the entropy function. This proves the theorem. □

3. Total Shannon entropy and entropy of continuous distribution

The definition (2.4) of entropy can be generalized straightforwardly to define the entropy of a discrete random variable.

Definition 3.1. Let $X \in \mathbb{R}$ denote a discrete random variable which takes on the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$, respectively, the entropy $H(X)$ of X is then defined by the expression [4]

$$H(X) = -k \sum_{i=1}^{n} p_i \log p_i. \tag{3.1}$$

Let us now generalize the above definition to take account of an additional uncertainty due to the observer himself, irrespective of the definition of random experiment. Let X denote a discrete random variable which takes the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$. We decompose the practical observation of X into two stages. First, we assume that $X \in L(x_i)$ with probability p_i, where $L(x_i)$ denotes the ith interval of the set $\{L(x_1), L(x_2), \ldots, L(x_n)\}$ of intervals indexed by x_i. The Shannon entropy of this experiment is $H(X)$. Second, given that X is known to be in the ith interval, we determine its exact position in $L(x_i)$ and we assume that the entropy of this experiment is $U(x_i)$. Then the global entropy associated with the random variable X is given by

$$H_T(X) = H(X) + \sum_{i=1}^{n} p_i U(x_i). \tag{3.2}$$

Let h_i denote the length of the ith interval $L(x_i)$, $(i = 1, 2, \ldots, n)$, and define

$$U(x_i) = k \log h_i. \tag{3.3}$$

We have then

$$H_T(X) = H(X) + k \sum_{i=1}^{n} p_i \log h_i = -k \sum_{i=1}^{n} p_i \log \frac{p_i}{h_i}. \tag{3.4}$$

The expression $H_T(X)$ given by (3.4) will be referred to as the total entropy of the random variable X. The above derivation is physical. In fact, what we have used is merely a randomization of the individual event $X = x_i$ $(i = 1, 2, \ldots, n)$ to account for the additional uncertainty due to the observer himself, irrespective of the definition of random experiment [4]. We will derive the expression (3.4) axiomatically as generalization of Theorem 2.1.

THEOREM 3.2. *Let the generalized entropy (2.3) satisfy, in addition to Axioms 1 to 4 of Theorem 2.1, the boundary conditions*

$$\phi_i(1) = k\log h_i \quad (i = 1, 2, \ldots, n) \tag{3.5}$$

to take account of the postobservational uncertainty, where h_i is the length of the ith class $L(x_i)$ (or width of the observational value x_i). Then the entropy function reduces to the form of the total entropy (3.4).

Proof. The procedure is the same as that of Theorem 2.1 up to the relation (2.13):

$$\phi'(p_j) = A\log p_j + B. \tag{3.6}$$

Integrating (3.6) with respect to p_j and using the boundary condition (3.5), we have

$$\phi(p_j) - k\log h_j = Ap_j\log p_j + (B - A)p_j - B \tag{3.7}$$

so that the generalized entropy (2.3) reduces to the form

$$\sum_j \phi(p_j) = -k\sum_{j=1}^{n} p_j\log\frac{p_j}{h_j}, \tag{3.8}$$

where we have taken $A = -k < 0$ for the same unit of measurement of entropy and the negative sign to take account of **Axiom 1**. The constants appearing in (3.8) have been neglected without any loss of characteristic properties. The expression (3.8) is the required expression of total entropy obtained earlier. □

Let us now see how to obtain the entropy of a continuous probability distribution as a limiting value of the total entropy $H_T(X)$ defined above. For this let us first define the differential entropy $H(X)$ of a continuous random variable X.

Definition 3.3. The differential entropy $H_C(X)$ of a continuous random variable with probability density $f(x)$ is defined by [2]

$$H_C(X) = -k\int_R f(x)\log f(x)dx, \tag{3.9}$$

where R is the support set of the random variable X. We divide the range of X into bins of length (or width) h. Let us assume that the density $f(x)$ is continuous within the bins. Then by mean-value theorem, there exists a value x_i within each bin such that

$$hf(x_i) = \int_{ih}^{(i+1)h} f(x)dx. \tag{3.10}$$

We define the quantized or discrete probability distribution $(p_1, p_2, \ldots, p_n)$ by

$$p_i = \int_{ih}^{(i+1)h} f(x)dx \tag{3.11}$$

so that we have then

$$p_i = hf(x_i). \tag{3.12}$$

The total entropy $H_T(X)$ defined for $h_i = h$ $(i = 1, 2, \ldots, n)$,

$$H_T(X) = -k \sum_{i=1}^{n} p_i \log \frac{p_i}{h}, \tag{3.13}$$

then reduces to the form

$$H_T(X) = -k \sum_{i=1}^{n} hf(x_i) \log f(x_i). \tag{3.14}$$

Let $h \to 0$, then by definition of Riemann integral, we have $H_T(X) \to H(X)$ as $h \to 0$, that is,

$$\lim_{h \to 0} H_T(X) = H_C(X) = -k \int_R f(x) \log f(x) dx. \tag{3.15}$$

Thus we have the following theorem.

THEOREM 3.4. *The total entropy $H_T(X)$ defined by (3.13) approaches to the differential entropy $H_C(X)$ in the limiting case when the length of each bin tends to zero.*

4. Application: differential entropy and entropy in classical statistics

The above analysis leads to an important relation connecting quantized entropy and differential entropy. From (3.13) and (3.15), we see that

$$-k \sum_{i=1}^{n} p_i \ln p_i \longrightarrow -k \int_R f(x) \ln \{hf(x)\} dx \tag{4.1}$$

showing that when $h \to 0$ that is, when the length of the bins h is very small, the quantized entropy given by the left-hand side of (4.1) approaches not to the differential entropy $H_C(X)$ defined in (3.9) but to the form given by the right-hand side of (4.1) which we call modified differential entropy. This relation has important physical significance in statistical mechanics. As an application of this relation, we now find the expression of classical entropy as a limiting case of quantized entropy.

Let us consider an isolated system with configuration space volume V and a fixed number of particles N, which is constrained to the energy shell $R = (E, E + \Delta E)$. We consider the energy shell rather than just the energy surface because the Heisenburg uncertainty principle tells us that we can never determine the energy E exactly. we can make ΔE as small as we like. Let $f(X^N)$ be the probability density of microstates defined on the phase space $\Gamma = \{X^N = (q_1, q_2, \ldots, q_{2N}; p_1, p_2, \ldots, p_{2N})\}$. The normalized condition is

$$\int_R f(X^N) X^N = 1, \tag{4.2}$$

where

$$R = \{X^N : E < H(X^N) < E + \Delta E\}. \tag{4.3}$$

Following (4.1), we define the entropy of the system as

$$S = -k \int f(X^N) \ln \{C^N f(X^N)\} dX^N. \tag{4.4}$$

The constant C^N appearing in (4.4) is to be determined later on. The probability density for statistical equilibrium determined by maximizing the entropy (4.4) subject to the condition (4.2) leads to

$$
\begin{aligned}
f(X^N) &= \frac{1}{\Omega(E, V, N)} \quad \text{for } E < H(X^N) < E + \Delta E \\
&= 0 \quad \text{otherwise,}
\end{aligned}
\tag{4.5}
$$

where $H(X^N)$ is the Hamiltonian of the system, $\Omega(E, V, N)$ is the volume of the energy shell $(E, E + \Delta E)$ [10]. Putting (4.5) in (4.4), we obtain the entropy of the system as [10]

$$S = k \ln \left\{ \frac{\Omega(E, V, N)}{C^N} \right\}. \tag{4.6}$$

The constant C^N has the same unit as $\Omega(E, V, N)$ and cannot be determined classically. However, it can be determined from quantum mechanics. Then we have $C^N = (h)^{3N}$ for distinguishable particles and $C^N = N!(h)^{3N}$ for indistinguishable particles. From Heisenberg uncertainty principle, we know that if h is the volume of a single state in phase space, then $\Omega(E, V, N)/(h)^{3N}$ is the total number of microstates in the energy shell $(E, E + \Delta E)$. The expression (4.6) then becomes identical to the Boltzmann entropy. With this interpretation of the constant C^N, the correct expression of classical entropy is given by [6, 10]

$$S = -k \int_R f(X^N) \ln \{(h)^{3N} f(X^N)\} dX^N. \tag{4.7}$$

The classical entropy that follows a limiting case of von Neumann entropy is given by [14]

$$S_d = -k \int_R \frac{f(X^N)}{(h)^{3N}} \ln \{f(X^N)\} dX^N. \tag{4.8}$$

This is, however, different from the one given by (4.7) and it does not lead to the form of Boltzmann entropy (4.6).

5. Conclusion

The literature on the axiomatic derivation of Shannon entropy is vast [1, 8]. The present approach is, however, different. This is based mainly on the postulates of additivity and concavity of entropy function. These are, in fact, variant forms of additivity and nondecreasing characters of entropy in thermodynamics. The concept of additivity is dormant in many axiomatic derivations of Shannon entropy. It plays a vital role in the foundation

of Shannon information theory [15]. Nonadditive entropies like Renyi entropy and Tsallis entropy need a different formulation and lead to different physical phenomena [11, 13]. In the present paper, we have also provided a new axiomatic derivation of Shannon total entropy which in the limiting case reduces to the expression of modified differential entropy (4.1). The modified differential entropy together with quantum uncertainty relation provides a mathematically strong approach to the derivation of the expression of classical entropy.

References

[1]   J. Aczél and Z. Daróczy, *On Measures of Information and Their Characterizations*, Mathematics in Science and Engineering, vol. 115, Academic Press, New York, 1975.

[2]   T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley & Sons, New York, 1991.

[3]   E. T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. (2) **106** (1957), 620–630.

[4]   G. Jumarie, *Relative Information. Theories and Applications*, Springer Series in Synergetics, vol. 47, Springer, Berlin, 1990.

[5]   J. N. Kapur, *Measures of Information and Their Applications*, John Wiley & Sons, New York, 1994.

[6]   L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon Press, Oxford, 1969.

[7]   V. Majernik, *Elementary Theory of Organization*, Palacky University Press, Olomouc, 2001.

[8]   A. Mathai and R. N. Rathie, *Information Theory and Statistics*, Wiley Eastern, New Delhi, 1974.

[9]   D. Morales, L. Pardo, and I. Vajda, *Uncertainty of discrete stochastic systems: general theory and statistical inference*, IEEE Trans. Syst., Man, Cybern. A **26** (1996), no. 6, 681–697.

[10]   L. E. Reichl, *A Modern Course in Statistical Physics*, University of Texas Press, Texas, 1980.

[11]   A. Rényi, *Probability Theory*, North-Holland Publishing, Amsterdam, 1970.

[12]   C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press, Illinois, 1949.

[13]   C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Statist. Phys. **52** (1988), no. 1-2, 479–487.

[14]   A. Wehrl, *On the relation between classical and quantum-mechanical entropy*, Rep. Math. Phys. **16** (1979), no. 3, 353–358.

[15]   T. Yamano, *A possible extension of Shannon's information theory*, Entropy **3** (2001), no. 4, 280–292.

C. G. Chakrabarti: Department of Applied Mathematics, University of Calcutta, Kolkata 700009, India

E-mail address: cgcappmath@caluniv.ac.in

Indranil Chakrabarty: Department of Mathematics, Heritage Institute of Technology, Chowbaga Road, Anandapur, Kolkata 700107, India

E-mail address: indranilc@indiainfo.com