

DERIVATION OF SOME NEW DISTRIBUTIONS IN STATISTICAL MECHANICS USING MAXIMUM ENTROPY APPROACH

Amritansu RAY

*Department of Mathematics, Rajyadharpur Deshbandhu Vidyapith, Serampore, Hooghly
– 712203, West Bengal, India.*

amritansu_ray06@yahoo.co.in

S.K. MAJUMDER

*Department of Mathematics, Bengal Engineering and Science University (BESU),
Shibpur, Howrah – 711103, West Bengal, India.*

majumder_sk@yahoo.co.in

Received: September 2012 / Accepted: June 2013

Abstract: The maximum entropy principle has been earlier used to derive the Bose Einstein(B.E.), Fermi Dirac(F.D.) & Intermediate Statistics(I.S.) distribution of statistical mechanics. The central idea of these distributions is to predict the distribution of the microstates, which are the particle of the system, on the basis of the knowledge of some macroscopic data. The latter information is specified in the form of some simple moment constraints. One distribution differs from the other in the way in which the constraints are specified. In the present paper, we have derived some new distributions similar to B.E., F.D. distributions of statistical mechanics by using maximum entropy principle. Some proofs of B.E. & F.D. distributions are shown, and at the end some new results are discussed.

Keywords: Bose-Einstein (B.E.) distribution, Fermi-Dirac (F.D.) distribution, Lagrange's multiplier, Shannons' measure, Jaynes principle.

MSC: 94A17, 82B10.

1. INTRODUCTION

The term “entropy” was introduced by Clausius in nineteenth-century thermodynamics, and is the subject of second law of thermodynamics, which states that in an isolated thermodynamic system, entropy will either remain constant or increase towards its maximum but cannot decrease. As we know, an isolated system is the one which is closed to inputs of both matter and energy; so in an isolated system, the system will gradually become more and more disordered, until it reaches maximum entropy. This means that since no new heat energy can be added, the system can never become hotter, but can only maintain the same temperature, or become colder. As it loses heat over time, its entropy increases, until finally it reaches maximum. This state of maximum entropy is called thermodynamic equilibrium. Such thermodynamic system are “irreversible system” where heat cannot flow from colder to hotter parts of the system, but only from hotter to colder areas. Thermodynamic entropy is denoted by the symbol S , and the formula for change in entropy is:

$dS = \frac{dQ}{T}$, where S is entropy, Q is heat, and T is temperature of the system. The difference in two entropy states S_1, S_2 is:

$S_2 - S_1 = \int_1^2 \frac{dQ}{T}$ (Irreversible process), when the system is in thermodynamic equilibrium then $dS = 0$.

In physics, work and entropy are inversely related. The principal way to decrease entropy is to do work through the expenditure of free energy. If free energy is available and expended to do work then, the system becomes more orderly and entropy decreases. But if all available energy has been expended, hence no more work is affordable; entropy will either remain constant or increase.

Boltzmann’s Entropy

In addition to thermodynamic (or heat-change) entropy, physicists also study entropy statistically. The statistical or probabilistic study of entropy is presented in Boltzmann’s law. Boltzmann’s equation is somewhat different from the original Clausius (thermodynamic) formulation of entropy. First, the Boltzmann formulation is structured in terms of probabilities, while the thermodynamic formulation does not entail the calculation of probabilities. The thermodynamic formulation can be characterized as a mathematical formulation, while the Boltzmann formulation is statistical. Second, the Boltzmann equation yields a value of entropy (S) while the thermodynamic formulation yields only a value for the change in entropy (dS). Third, there is a shift in content, as the Boltzmann equation was developed for research on gas molecules, rather than thermodynamics. Fourth, by incorporating probabilities, the Boltzmann equation focuses on micro-states, and thus explicitly introduces the question of the relationship between macro-states and microstates.

Boltzmann investigated such micro-states and defined entropy in a new way, such that the macroscopic maximum- entropy state corresponded to a thermodynamic configuration, which could be realized by the maximum number of different micro-states.

He noticed that entropy of a system can be considered as a measure of the disorder in the system, and if a system has many degrees of freedom, the number measuring the degree of disorder also measures the uncertainty, in a probabilistic sense, of the particular micro-states.

We consider a system with N micro-states $N_i (i = 1, 2, \dots, n)$. Then the statistical Disorder is given by [Carnop, 1977] $W = \frac{N!}{N_1! N_2! \dots N_n!}$. Let us now take an approximate value of W for large $N!$ Using Stirling's approximation, we have

$$S = -K \sum_{i=1}^n p_i \ln p_i, \text{ where } p_i = \frac{N_i}{N} \text{ is the probability of the occurrence of micro states } i;$$

S is the entropy and K is a positive constant called Boltzmann's constant. Boltzmann was the first to emphasize the probabilistic meaning of entropy, and the probabilistic nature of thermodynamics.

Entropy in Information Theory

Unlike the first two entropy approaches (thermodynamic entropy and Boltzmann's entropy), the third major form of entropy was not developed within the field of physics but in a new field known as information theory (also known as communication theory). A fundamental step in using entropy in the new contexts unrelated to thermodynamics was provided by Shannon (1948), who realized that entropy could be used to measure types of disorder other than those of thermodynamic micro-states. Shannon was interested in information theory, particularly the ways information can be conveyed via a message. This led him to examine probability distributions in a very general sense, and to find some ways to measure the level of uncertainty in different distributions. For example, suppose that probability distribution for the outcome of a coin-toss experiment is $P(H) = 0.999$ and $P(T) = 0.001$. One is likely to notice that there is much more "certainty" than "uncertainty" about the outcome of this experiment and the probability distribution, too. If, on the other hand, the probability distribution governing that same experiment were $P(H) = 0.5$ and $P(T) = 0.5$, then there is much less "certainty" and much more "uncertainty", when compared with the previous distribution. But how to quantify this uncertainty? Is there any algebraic function that measures the amount of uncertainty in any probabilistic distribution, in terms of the individual probabilities?

From these and other types of simple examples, Shannon was able to devise a set of criteria that any measure of uncertainty may satisfy. He then tried to find an algebraic form which could satisfy his criteria, and discovered that there was only one formula that fitted. The amount of uncertainty in any discrete probability distribution

where $p_i, i = 1, 2, \dots, n$, is the probability of an event, is proportional to the $-\sum_{i=1}^n p_i \ln p_i$.

This is identical to the previous entropy relation if the constant of probability is taken as Boltzmann constant K . Thus, Shannon showed that entropy that measures the amount of disorder in a thermodynamic system, also measures the amount of uncertainty in any probability distribution. Let us now give the formal definition of Shannon Entropy: Consider a random experiment $\alpha = (p_1, p_2, \dots, p_n)$ whose possible outcomes

have probabilities $p_i, i = 1, 2, \dots, n$, which are known. Can we predict the outcome obtained? As a matter of fact, the degree of difficulty of this prediction will vary with the definition of α and is related to the uncertainty included in α . So the question is then the following: can we measure the amount of uncertainty? We shall denote such uncertainty measure by $H(\alpha) = H(p_1, p_2, \dots, p_n)$. The most common, as well as the most useful measure of uncertainty is the Shannon's informational entropy (which should satisfy some basic requirements), defined as follows:

Definition 1: Let (p_1, p_2, \dots, p_n) be the probability of occurrence of the events E_1, E_2, \dots, E_n associated with a random experiment. The Shannon Entropy probability distribution (p_1, p_2, \dots, p_n) of the random experiment system α is defined by

$$H(\alpha) = H(p_1, p_2, \dots, p_n) = -K \sum_{i=1}^n p_i \ln p_i .$$

Where $0 \ln 0 = 0$, and K is the unit of measurement of entropy. The above definition is generalized straight forwardly from the definition of entropy of a random variable.

Definition 2: Let $X \in R$ be discrete random variable which takes a value x ($i=1, 2, \dots, n$) with probability $p_i, i = 1, 2, \dots, n$, then the entropy $H(X)$ of X is defined

$$\text{by the expression } H(X) = -K \sum_{i=1}^n p_i \ln p_i .$$

Examination of H reveals that the measure varies from a minimum value of zero to a maximum value of $\log K$ (where K is the number of categories). If a sample has N objects (such as persons), the minimum value of zero is attained when all N objects occupy a single category. The maximum value of H ($\log K$) is attained when all N objects are evenly distributed among the K categories, so that each category contains exactly $1/N$ of the total sample. There are a few points to note about the H measure, and its comparisons with Boltzmann's entropy. The maximum value of H is not a fixed value, but is dependent upon the number of categories, K . What degree of problem it makes, depends upon the particular variable or subject matter to which K is applied. For example, if one wishes to compute H for social class, then there may be disagreement, for example, about whether the social class comprises three categories (lower class, middle class, and upper class), or five categories (lower class, lower-middle class, middle class, upper-middle class, and upper class). While minimum H is fixed at zero in both cases, maximum H varies, depending on the number of categories chosen, be them three or five. Another point to note when comparing H with Boltzmann's equation is that the formula for H specifically includes both population or sample size (N) and category size (K), and that both of these are used in computing the probabilities p_i . In contrast, Boltzmann's equation does not identify the population size and categories, instead, it is written in terms of probability of an event occurring all the ways possible. These represent two different ways of approaching the macro-state/microstate problem.

A further issue with H is the question of how it should be interpreted. Consider the application of H to a situation where $K = 5$, as in the case of the social class given

above. Here, H can vary from a minimum of zero (when everyone is in the same income category) to a maximum of $\log K$ (when each of the K categories contains exactly $1/K$ of the sample). Statistically, this maximum state is the “most probable state,” or most likely state. It is also the least predictable state and thus, the one that exhibits the most uncertainty or disorder.

That is, if one wishes to predict the social class of an individual, the prediction is certain for the case where H equals zero, as all persons have the same income. Thus, when H is a minimum, there is no uncertainty, and the social class is perfectly predictable. However, when H is maximum, then the probability of assignment to each category is basically random, resulting in an even distribution of individuals across class categories. This means that uncertainty is maximized, and it is difficult to predict a person’s social class. Since H was developed within information theory, it is often referred to as a measure of information. The main idea here is that minimum H yields minimum uncertainty (high information), while maximum H represents maximum uncertainty (low information). Thus, by calling H information, scholars must face the uncomfortable anomaly or irony that minimum information yields maximum certainty, and that maximum information yields minimum certainty. If one is comfortable with an equation of information and uncertainty, then H can be termed as a measure of information. If one is more comfortable with H as a direct measure of uncertainty or disorder (it clearly is), then it is better to term H a measure of entropy. In contrasting H with Clausius’ thermodynamic entropy and Boltzmann’s entropy, several differences are apparent. Thermodynamic entropy is defined solely in terms of heat and temperature, as shown earlier. In contrast, Boltzmann’s entropy is written in terms of the behavior of gas molecules. H is a more open or generic measure, which is more content free, and can be applied to any set of discrete categories for which data exists. Shannon’s H is strikingly similar to Boltzmann’s entropy in terms of its equation. Since Boltzmann’s entropy and Shannon’s H are statistical formulations presented in terms of probabilities, and since the formulas are so similar, many statisticians may have little difficulty making the transition from Boltzmann’s entropy to Shannon’s H (or vice versa). However, students of Shannon’s H may have more difficult time understanding thermodynamic entropy, and resolving differences between thermodynamic entropy and Shannon’s H . Similarly, students whose first introduction to entropy was in thermodynamics, may not be sure that Boltzmann’s entropy is “real” entropy, but may accept it as a valid physics entity. However, they may consider this statistical computation to be less rigorous than the original thermodynamic entropy. These scholars may have real difficulty in accepting Shannon’s H as an entropy measure. They may consider it a valid information- theory measure, or entropy analog, but may not consider it a valid entropy measure to the same degree that they would accept thermodynamic entropy or even Boltzmann’s entropy.

The maximum entropy principle of Jaynes [7] has been frequently used to derive the distribution of statistical mechanics by maximizing the entropy of the system subject to some given constraints.

The Maxwell-Boltzmann distribution is obtained when there is only one constraint on system which prescribes the expected energy per particle of the system by J.N. Kapur [2].

Bose-Einstein (B.E.) distribution, Fermi-Dirac (F.D.) distribution, and Intermediate statistics (I.S.) distributions are obtained by maximizing the entropy subject

to two constraints by Forte & Sempi [1], by J.N.Kapur [2,3] and Kapur & Kesavan [4,5] which are

- (i) The expected number of particle of the system is prescribed,
- (ii) The expected total energy of the system is prescribed, respectively.

Though these distributions arose in the first instance in statistical mechanics, they are widely applicable in urban & regional planning, transportation studies (S.K.Mazumder [6]), finance, banking, and economics (Kapur [2], Kapur Kesavan [4,5]).

In the present paper, we obtain such distributions when only one constraint is given for B.E. & F.D. distribution i.e. we have derived some new distributions, relaxing the assumptions usually used when B.E. and F.D. distributions are derived.

2. DISCUSSION

Let p_1, p_2, \dots, p_n be the probabilities of a particle having energy levels $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, respectively, and let the expected value of energy be prescribed as $\bar{\varepsilon}$, and then to get the maximum entropy probability distribution (MEPD) of Maxwell-Boltzmann, we maximize the Shannon's measure of entropy

$$-\sum_{i=1}^n p_i \ln p_i \quad (1)$$

Subject to

$$\sum_{i=1}^n p_i = 1 \quad (2)$$

$$\sum_{i=1}^n p_i \varepsilon_i = \bar{\varepsilon} \quad (3)$$

Let the Lagrangian be

$$L = -\sum_{i=1}^n p_i \ln p_i - (\lambda - 1) \left[\sum_{i=1}^n p_i - 1 \right] - \mu \left[\sum_{i=1}^n p_i \varepsilon_i - \bar{\varepsilon} \right] \quad (4)$$

differentiating with respect to p_i 's, we get

$$\begin{aligned} \ln p_i + \lambda + \mu \varepsilon_i &= 0 \\ \text{or, } p_i &= \exp(-\lambda - \mu \varepsilon_i) \end{aligned}$$

where λ, μ are to be determined by using (2),(3) so that

$$p_i = \frac{\exp(-\mu\varepsilon_i)}{\sum_{i=1}^n \exp(-\mu\varepsilon_i)}, i = 1, 2, \dots, n, \tag{5}$$

where $\frac{\sum_{i=1}^n \varepsilon_i \exp(-\mu\varepsilon_i)}{\sum_{i=1}^n \exp(-\mu\varepsilon_i)} = \bar{\varepsilon}$, (6)

and $n =$ number of energy levels, $\mu = \frac{1}{kT}$, k is Boltzmann constant, and T is absolute temperature.

Equation (5) is the well-known Maxwell-Boltzmann distribution.

(1)The new MEPD when we take the B.E. measure of entropy is

$$-\sum_{i=1}^n p_i \ln p_i + \sum_{i=1}^n (1 + p_i) \ln(1 + p_i), \text{ subject to (2),(3).}$$

Here, we show the MEPD given by

$$p_i = \frac{1}{\exp(\mu + v\varepsilon_i) - 1}. \tag{7}$$

Let us consider the Lagrangian

$$L = -\sum_{i=1}^n p_i \ln p_i + \sum_{i=1}^n (1 + p_i) \ln(1 + p_i) - \mu(\sum_{i=1}^n p_i - 1) - v(\sum_{i=1}^n p_i \varepsilon_i - \bar{\varepsilon}), \tag{8}$$

maximizing for variations in p_i 's, we get

$$-(\ln p_i + 1) + \{\ln(1 + p_i) + 1\} - \mu - v\varepsilon_i = 0 \tag{9}$$

or, $-\ln p_i + \ln(1 + p_i) = \mu + v\varepsilon_i$

$$\text{or, } \frac{1 + p_i}{p_i} = \exp(\mu + v\varepsilon_i) \tag{10}$$

$$\text{or, } p_i = \frac{1}{\exp(\mu + v\varepsilon_i) - 1} \tag{11}$$

(2) In the second case, we show the probability distributions which maximize the entropy measure

$$-\sum_{i=1}^n p_i \ln p_i - \sum_{i=1}^n (1-p_i) \ln(1-p_i). \quad (12)$$

Subject to (2) & (3) is given by

$$p_i = \frac{1}{\exp(\mu + v\varepsilon_i) + 1} \quad (13)$$

The above can be proved in similar manner by choosing the Lagrangian

$$L \equiv -\sum_{i=1}^n p_i \ln p_i - \sum_{i=1}^n (1-p_i) \ln(1-p_i) - \mu \left(\sum_{i=1}^n p_i - 1 \right) - v \left(\sum_{i=1}^n p_i \varepsilon_i - \bar{\varepsilon} \right). \quad (14)$$

Here, maximizing for variations in p_i , we get

$$-(\ln p_i + 1) - \{-\ln(1-p_i) - 1\} - \mu - v\varepsilon_i = 0 \quad (15)$$

$$\text{or, } -\ln p_i + \ln(1-p_i) = \mu + v\varepsilon_i \quad (16)$$

$$\text{or, } \frac{1-p_i}{p_i} = \exp(\mu + v\varepsilon_i) \quad (17)$$

$$\therefore p_i = \frac{1}{\exp(\mu + v\varepsilon_i) + 1} \quad (18)$$

The reverse cases

$$(3) \text{ If } p_i = [\exp(\mu + v\varepsilon_i) + 1]^{-1} \text{ maximizes } \sum_{i=1}^n f(p_i) \quad (19)$$

Subject to (2), (3)

$$\text{then, } f(x) = -x \ln x - (1-x) \ln(1-x) \quad (20)$$

Let the Langrangian

$$L \equiv \sum_{i=1}^n f(p_i) - \mu \left(\sum_{i=1}^n p_i - 1 \right) - v \left(\sum_{i=1}^n p_i \varepsilon_i - \bar{\varepsilon} \right) \quad (21)$$

$$\therefore \frac{\partial L}{\partial p_i} = 0 \text{ gives us } f'(p_i) - \mu - v\varepsilon_i = 0 \quad (22)$$

$$\therefore f'(p_i) = \mu + v\varepsilon_i$$

$$\text{or, } f'((\exp(\mu + v\varepsilon_i) + 1)^{-1}) = \mu + v\varepsilon_i \quad (23)$$

$$\text{or, } f'\left(\frac{1}{e^z + 1}\right) = \ln(e^z + 1) \text{ where, } z = \mu + v\varepsilon_i \quad (24)$$

$$\text{or, } f'(x) = \ln\left(\frac{1}{x} - 1\right) = \ln(1-x) = \ln(1-x) - \ln x \tag{25}$$

Integrating, we get

$$f(x) = -x \ln x - (1-x) \ln x . \tag{26}$$

Without loss of generality, the constant of integration is assumed to be zero.

$$(4) \text{ Similarly if } p_i = [\exp(\mu + \nu \varepsilon_i) - 1]^{-1} , \tag{27}$$

maximizes $\sum_{i=1}^n f(p_i)$ subject to (19) & (20),

$$\text{then } f(x) = -x \ln x + (1+x) \ln(1+x) . \tag{27}$$

Progressing in the same manner, we get

$$f'\left(\frac{1}{e^z - 1}\right) = \ln(e^z - 1 + 1)$$

$$\text{or, } f'(x) = \ln\left(\frac{1}{x} + 1\right) \tag{29}$$

$$\text{or, } f'(x) = \ln(1+x) - \ln x \tag{30}$$

$$\text{or, } f(x) = -x \ln x + (1+x) \ln(1+x) \tag{31}$$

A new result:

(5) Let p_{ij} be the probability of there being j particle in the i^{th} energy level. Let q_i be a priori probability, assumed as known, of the system being in the i^{th} energy level, so that $\sum_{i=1}^n q_i = 1$. (32)

Since the number of particles in the i^{th} energy level can be 0 and 1 instead of the values 0, 1, 2, 3, we get,

$$\sum_{j=0}^1 p_{ij} = 1, i = 1, 2, 3, \dots, n . \tag{33}$$

Let the expected numbers of particles and the expected energy to be prescribed are a and b , respectively, so that

$$\sum_{i=1}^n q_i \sum_{j=0}^1 j p_{ij} = a \tag{34}$$

$$\text{and } \sum_{i=1}^n q_i \varepsilon_i \sum_{j=0}^1 j p_{ij} = b \quad (35)$$

$$\text{then, we know } p_{ij} = a_i \exp(-(\mu + v\varepsilon_i)j), \quad (36)$$

$$\text{and get } a_i = 1 / (\exp(-(\mu + v\varepsilon_i)) + 1). \quad (37)$$

Now if \bar{n}_i is the expected number of particle in the i^{th} energy level, by using (36) and (37), we have

$$\bar{n}_i = \sum_{j=0}^1 j p_{ij} = \frac{1}{\exp(\mu + v\varepsilon_i) + 1}. \quad (38)$$

Now, we prove that for a F.D. distribution

$$s_{\max} = -\sum_{i=1}^n \bar{n}_i \ln \bar{n}_i - \sum_{i=1}^n (1 - \bar{n}_i) \ln(1 - \bar{n}_i)$$

by using (33), (34), (35), (36), (37) and (38),

the maximum entropy S given by

$$\begin{aligned} s_{\max} &= -\sum_{i=1}^n q_i \sum_{j=0}^1 p_{ij} \ln p_{ij} \\ &= -\sum_{i=1}^n q_i \sum_{j=0}^1 p_{ij} \ln [a_i \exp(-(\mu + v\varepsilon_i)j)] \\ &= -\sum_{i=1}^n q_i \left[\sum p_{ij} \ln a_i - \mu \sum j p_{ij} - v \sum j \varepsilon_i p_{ij} \right] \\ &= -\sum q_i \ln a_i + \mu a + vb \\ &= -\sum_{i=1}^n q_i \ln [1 / (\exp(-(\mu + v\varepsilon_i)) + 1)] + \sum (\mu + v\varepsilon_i) q_i \sum j p_{ij} \\ &= \sum q_i \ln [1 + \exp(-(\mu + v\varepsilon_i))] + (\mu + v\varepsilon_i) \bar{n}_i \\ &= \sum q_i \left\{ \ln \left(1 + \frac{\bar{n}_i}{1 - \bar{n}_i} \right) + \bar{n}_i \ln \left(\frac{1}{\bar{n}_i} - 1 \right) \right\} \\ &= \sum_{i=1}^n q_i \left\{ \ln \left(\frac{1}{1 - \bar{n}_i} \right) + \bar{n}_i \ln \left(\frac{1 - \bar{n}_i}{\bar{n}_i} \right) \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n q_i \left\{ -\ln(1 - \bar{n}_i) + \bar{n}_i \ln(1 - \bar{n}_i) - \bar{n}_i \ln \bar{n}_i \right\} \\
&= \sum_{i=1}^n q_i \left\{ -\bar{n}_i \ln \bar{n}_i - (1 - \bar{n}_i) \ln(1 - \bar{n}_i) \right\} \tag{39}
\end{aligned}$$

3. CONCLUSION

The measures discussed in this paper are closely related to the classical distributions, seen in quantum mechanics of a gas of very weakly coupled particles where the total wave function is either anti symmetric or symmetric. In the anti symmetric case, we are led to F.D. where any given elementary state can be occupied by at most one particle. In the symmetric case, we get B.E. where there is no restriction on the occupation number of a given particle. Regarding the applicability of Bose Einstein and Fermi-Dirac entropy, we state that though Shannon Entropy is of wide range of applicability, B.E. and F.D. distributions have been applied successfully in case of work trip distribution and commodity distribution, respectively [8].

4. REFERENCE

- [1] B., Forte and C. Sempì, "Maximizing Conditional Entropy : A derivation of quantal statistics", *Rendi Conta de Mathematics*, 9 (1976) 551-556.
- [2] J.N., Kapur, *Maximum Entropy Models in Science and Engineering*, Wiley Eastern, New Delhi and John Wiley, New York, 1980.
- [3] J.N., Kapur, "Non-additive measures of entropy and distributions of statistical mechanics", *Ind. Jour. Pure App. Math*, 14 (11) (1977) 1372-1384.
- [4] J.N., Kapur & H.K., Kesavan, *The Generalized maximum entropy principle with Application*, Sandford Educational Press. University of Waterloo, Canada, 1989.
- [5] H.K., Kesavan & J.N. Kapur, *Entropy Optimization Principle & their Application*, Academic Press, New York, 1992.
- [6] S.K., Mazumder, "Maximum Entropy and utility in a transportation system", *Yugoslav Journal of Operations Research*, 9 (1) (1999) 27-34.
- [7] E.T., Jaynes, "Information theory and Statistical Mechanics", *Raynal Review*, 106 (1957) 620-630.
- [8] C., Ficks, "Entropy and information theory", *Env and Planning*, 14A (1985) 679.