MB DISTRIBUTION AND ITS APPLICATION USING MAXIMUM ENTROPY APPROACH

Suman BHADRA
Research Scholar
Department of Mathematics
IIEST, Shibpur (Formerly BESU, Shibpur)
sb.math@gmail.com
Sanat K. MAJUMDER
Professor, Department of Mathematics
IIEST, Shibpur (Formerly BESU, Shibpur)
majumder_sk@yahoo.co.in

Abstract: Maxwell Boltzmann distribution with maximum entropy approach has been used to study the variation of political temperature and heat in a locality. We have observed that the political temperature rises without generating any political heat when political parties increase their attractiveness by intense publicity, but voters do not shift their loyalties. It has also been shown that political heat is generated and political entropy increases with political temperature remaining constant when parties do not change their attractiveness, but voters shift their loyalties (to more attractive parties).

1. INTRODUCTION

1.1. Entropy in thermodynamics

The first law of thermodynamics gives a generalization of the law of conservation of energy including heat energy according to which the total energy of the universe remains constant. Every system left to itself changes spontaneously at a final state of equilibrium. As a well known example is the passage of heat from a hotter region to a colder until both regions become of the same temperature, but if we think this process to occur in reverse order, it cannot happen spontaneously. Here comes the concept of the second law of thermodynamics, in addition to the first law of thermodynamics that determines the direction in which a process can take place in an isolated system, a system which is congested with inputs of both matter and energy. Two following statements form the basis of the second law of thermodynamics [1].
Kelvin Statement: There exists no thermodynamic transformation (A Thermodynamic transformation is a change of state) whose sole effect is to extract a quantity of heat from a given heat reservoir and to convert it entirely into work.
Clausius Statement: There exists no thermodynamic transformation whose sole effect is to extract a quantity of heat from a colder reservoir and to deliver it to a hotter reservoir.

The second law of thermodynamics enables us to define a state function **S,** entropy. The term "entropy" was introduced in thermodynamics by Clausius in nineteenth-century [2] , and is the theme of the second law of thermodynamics, which states that in an isolated thermodynamic system, entropy cannot decrease but moreover, it will stay stable or augment towards its maximum. This is explained by Clausius Theorem.

Clausius Theorem: In any cyclic transformation through which the temperature is defined, the following inequality holds:

$$\oint \frac{dQ}{T} \leq 0$$

The integral extends over one cycle of transformation. The equality holds for reversible cyclic transformation.

In an isolated system, the system will steadily become more and more chaotic, 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 continue to have the same temperature or to become colder. As it loses heat over time, its entropy increases, awaiting finally to reach its maximum. This state of maximum entropy is called thermodynamic equilibrium which prevails when the thermodynamic state of the system does not change with time. Such thermodynamic systems are "irreversible transformations" where heat cannot flow from colder to hotter parts of the system, but only from hotter to colder areas.

1.2. Entropy in Information Theory

Information theory is a new branch of probability theory with extensive potential applications to communication systems. It was originated by scientists, while studying the statistical structure of electrical communication equipment. Information Theory has its origin in the twentieth century itself, when Hartley tried to develop a quantitative measure of information to assess the capabilities of the telecommunication systems. It is only during the last three decades or so, that this measure of information has been developed and its concept has found widespread use outside the telecommunication engineering. Mathematical theory of communication was principally originated by Claude Shannon in 1948 [3]. It plays an important role in modern communication theories, where a communication system is formulated as a stochastic or random process. The maximum entropy principle of Jaynes [4] has been frequently used to derive the distribution of statistical mechanics by maximizing the entropy of the system subject to some given constraints.

1.3. Entropy in MB distribution

In thermodynamics, a system of identical and distinguishable particles of any spin obey the MB Distribution law and we get information how a total fixed amount of energy is distributed among the various members of the aforementioned system in the most probable distribution. In Information Theory, Jaynes derived the distribution in a manner totally different from the classical derivations. The central idea of the distribution is to predict the distribution of the microstates which are the particles of the system on the basis of the knowledge of some macroscopic data. The macroscopic data are specified in the form of simple moment constraint. One distribution differs from another in the way in which the constraints are specified, and Maxwell-Boltzman distribution is obtained when there is only one constraint on the system that prescribes the expected energy per particle of the system [5], using Shannon entropy measure. 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, (see Forte and Sempi [6], J. N. Kapur [5, 7], and Kapur and Kesavan [8,9]). Though these distributions arose in the first instance in statistical mechanics (Amritansu Ray, S. K.

Majumder [10]), they are widely applicable in urban and regional planning, transportation studies (S.K.Mazumder [11]), finance, banking, and economics (Kapur [5], Kapur and Kesavan [8,9], W. Ximing[12]).

The paper is organised as follows: In Section 2, we describe our model; in Section 3 we derive the MB distribution function, followed by some results. In Section 4, we give the application of this MB Distribution in political entropy.

2. MODEL

In the present study we have considered a locality, where n political parties are there with different attractiveness. We find Maximum Entropy Probability Distribution (MEPD) for the proportion of voters voting for different political parties in this locality by using the concept of MB Distribution. Here, the input information of MB Distribution is the expected average attractiveness of the political parties in lieu of expected energy. The total proportion of voters is 1, which is the constraint of the distribution. Using this model, we will provide the concept of political entropy, political heat, and political temperature in this locality, which will serve as an isolated system without any transformation of voters and political parties from an external source.

3. RESULT AND DISCUSSION

3.1. The Maxwell – Boltzmann (MB) distribution:

Let $p_1, p_2, ..., p_n$ be the probabilities that a particle in a system has energies $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$, respectively. Suppose that the only information about the system is that the expected energy of the particles of the system is $\hat{\varepsilon}$, i.e. we are given the information

$$\sum_{i=1}^{n} p_i \varepsilon_i = \hat{\varepsilon}$$

$$\sum_{i=1}^{n} p_i = 1, p_i \geq 0, \forall i = 1, 2, ..., n$$

Now, applying Jaynes' Maximum Entropy principle we choose that probability distribution that maximizes $-\sum_{i=1}^{n} p_i \ln p_i$, subject to the above constraints, the Lagrangian is

$$L \equiv -\sum_{i=1}^{n} p_i \ln p_i - (\lambda - 1)(\sum_{i=1}^{n} p_i - 1) - \mu(\sum_{i=1}^{n} p_i \varepsilon_i - \hat{\varepsilon})$$

Setting the derivative of the Lagrangian with respect to $p_1, p_2, ..., p_n$ equal to zero, we get

$$-\ln p_i - p_i . \frac{1}{p_i} - (\lambda - 1) - \mu \varepsilon_i = 0 \qquad\qquad i = 1, 2, ..., n$$

$$\Rightarrow -\ln p_i - 1 - \lambda + 1 - \mu \varepsilon_i = 0 \qquad\qquad\qquad (1)$$

$$\Rightarrow -\ln p_i = \lambda + \mu \varepsilon_i$$

$$\Rightarrow \ln p_i = -\lambda - \mu \varepsilon_i$$

$$\Rightarrow p_i = e^{-\lambda - \mu \varepsilon_i} \qquad\qquad\qquad (2)$$

Applying the natural constraint, we get,

$$e^{\lambda} = \sum_{i=1}^{n} e^{-\mu\varepsilon_i} \tag{3}$$

Now from equation (1), we get,

$$p_i = \frac{e^{-\mu\varepsilon_i}}{\sum_{i=1}^{n} e^{-\mu\varepsilon_i}} \qquad i = 1, 2, ..., n$$

The discrete –variate distribution is called the Maxwell – Boltzmann distribution of statistical mechanics.

The Lagrange Multiplier μ is determined from,

$$f(\mu) = [\frac{\sum_{i=1}^{n} \varepsilon_i e^{-\mu\varepsilon_n}}{\sum_{i=1}^{n} e^{-\mu\varepsilon_i}}] - \hat{\varepsilon} = 0 \tag{4}$$

$$f'(\mu) = (\sum_{i=1}^{n} e^{-\mu\varepsilon_i})^{-2} [(\sum_{i=1}^{n} e^{-\mu\varepsilon_i})(-\sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu\varepsilon_i}) + (\sum_{i=1}^{n} \varepsilon_i e^{-\mu\varepsilon_i})^2]$$

$$= (\sum_{i=1}^{n} e^{-\mu\varepsilon_i})^{-2} [(\sum_{i=1}^{n} \varepsilon_i e^{-\mu\varepsilon_i})^2 - (\sum_{i=1}^{n} e^{-\mu\varepsilon_i})(\sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu\varepsilon_i})]$$

$f'(\mu) \leq 0$ by Cauchy – Schwarz's inequality, equality holds if and only if

$$\frac{a_1}{a_1(\varepsilon_1)^2} = \frac{a_2}{a_2(\varepsilon_2)^2} = ... = \frac{a_n}{a_n(\varepsilon_n)^2}$$

i.e., if and only if

$$\varepsilon_1 = \varepsilon_2 = ... = \varepsilon_n .$$

We shall assume that the energy levels are all different so that $f'(\mu) < 0$. So, $f(\mu)$ is a strictly decreasing function of μ. Without loss of generality, we can assume that

$\varepsilon_1 < \varepsilon_2 < .... < \varepsilon_n$

So that

$$a_1 > a_2 > a_3 > ... > a_n \qquad when\ \mu > 0$$

$$a_1 < a_2 < a_3 < ... < a_n \qquad when\ \mu < 0$$

When $\mu \to -\infty,\ a_n >> a_1, a_2, ..., a_{n-1}$

So that $f(-\infty) = \varepsilon_n - \hat{\varepsilon}$

Similarly, $f(\infty) = \varepsilon_1 - \hat{\varepsilon}$

$$f(0) = \frac{1}{n}(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n) - \hat{\varepsilon}$$

$$= \overline{\varepsilon} - \hat{\varepsilon}$$

If $\hat{\varepsilon} < \varepsilon_1$, then $f(\mu)$ will be positive throughout, and if $\hat{\varepsilon} > \varepsilon_n$, then $f(\mu)$ will be negative throughout. In either case, solution of $f(\mu)$ does not exist; $f(\mu)$ will have a solution only when $\varepsilon_1 < \overline{\varepsilon} < \varepsilon_n$.

3.2. Some results based on MB distribution

Result 1: $f(\mu) + \hat{\varepsilon}$ **is ratio of two convex functions of** μ .

From equation (2), we get

$$f(\mu) + \hat{\varepsilon} = \frac{\sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i}}{\sum_{i=1}^{n} e^{-\mu \varepsilon_i}}$$

$$= \frac{f_1(\mu)}{f_2(\mu)}$$

Now, we shall show that $f_1(\mu)$ and $f_2(\mu)$ are two convex functions of μ .

$$f_1(\mu) = \sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i}$$

$$\Rightarrow f_1'(\mu) = \sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i} (-\varepsilon_i)$$

$$\Rightarrow f_1'(\mu) = -\sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu \varepsilon_i}$$

$$\Rightarrow f_1''(\mu) = -\sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu \varepsilon_i} (-\varepsilon_i)$$

$$\Rightarrow f_1''(\mu) = \sum_{i=1}^{n} \varepsilon_i^3 e^{-\mu \varepsilon_i} > 0$$

So, $f_1(\mu)$ is a convex function of μ .

$$f_2(\mu) = \sum_{i=1}^{n} e^{-\mu \varepsilon_i}$$

$$\Rightarrow f_2'(\mu) = -\sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i}$$

$$\Rightarrow f_2''(\mu) = \sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu \varepsilon_i} > 0$$

So, $f_2(\mu)$ is a convex function of μ .

Result 2: $f'(-\infty) = 0, f'(\infty) = 0, f'(0) = -\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 + (\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i)^2$

From equation (2), we get

$$f(\mu) = \frac{\sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i}}{\sum_{i=1}^{n} e^{-\mu \varepsilon_i}} - \hat{\varepsilon}$$

$$f'(\mu) = (\sum_{i=1}^{n} e^{-\mu \varepsilon_i})^{-2}[(\sum_{i=1}^{n} e^{-\mu \varepsilon_i})(-\sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu \varepsilon_i}) + (\sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i})^2]$$

$$= (\sum_{i=1}^{n} e^{-\mu \varepsilon_i})^{-2}[(\sum_{i=1}^{n} \varepsilon_i e^{-\mu \varepsilon_i})^2 - (\sum_{i=1}^{n} e^{-\mu \varepsilon_i})(\sum_{i=1}^{n} \varepsilon_i^2 e^{-\mu \varepsilon_i})]$$

$$f'(0) = \frac{(\sum_{i=1}^{n} \varepsilon_i)^2 - n.\sum_{i=1}^{n} \varepsilon_i^2}{n^2}$$

$$= -\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 + (\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i)^2$$

(5)

Result 3: If $f'(\mu)$ vanishes at $\mu = 0$, it vanishes everywhere.

Using equation (3) and making it equal to 0, we get

$$-\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 + (\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i)^2 = 0$$

$$\Rightarrow \frac{1}{n^2}\sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2$$

$$\Rightarrow (n-1)\sum_{i=1}^{n} \varepsilon_i^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} \varepsilon_i^2 = 0$$

$$\Rightarrow \varepsilon_1 = \varepsilon_2 = ... = \varepsilon_n$$

i.e., $f'(\mu) = 0$ everywhere.

Result 4: For MB Distribution $\lambda_0 = \ln(\sum_{i=1}^{n} e^{-\mu \varepsilon_i})$ and $\lambda_1 = \mu$, then λ_0 is a convex function of λ_1.

As $\lambda_0 = \ln(\sum_{i=1}^{n} e^{-\mu \varepsilon_i})$

So,

$$\frac{d\lambda_0}{d\lambda_1} = \frac{1}{\sum_{i=1}^{n} e^{(-\lambda_1 \varepsilon_i)}} \sum_{i=1}^{n} (-\varepsilon_i)e^{(-\lambda_1 \varepsilon_i)}$$

$$\frac{d^2\lambda_0}{d\lambda_1^2} = \frac{\sum_{i=1}^{n} e^{(-\lambda_1 \varepsilon_i)} \sum_{i=1}^{n} \varepsilon_i^2 e^{(-\lambda_1 \varepsilon_i)} - \sum_{i=1}^{n}(-\varepsilon_i)e^{(-\lambda_1 \varepsilon_i)} \sum_{i=1}^{n}(-\varepsilon_i)e^{(-\lambda_1 \varepsilon_i)}}{(\sum_{i=1}^{n} e^{(-\lambda_1 \varepsilon_i)})^2}$$

$$\frac{d^2\lambda_0}{d\lambda_1^2} = \frac{\sum_{i=1}^{n} e^{(-\lambda_1 \varepsilon_i)} \sum_{i=1}^{n} \varepsilon_i^2 e^{(-\lambda_1 \varepsilon_i)} - \{\sum_{i=1}^{n} \varepsilon_i e^{(-\lambda_1 \varepsilon_i)}\}^2}{(\sum_{i=1}^{n} e^{(-\lambda_1 \varepsilon_i)})^2} \geq 0$$

Hence, λ_0 is a convex function of λ_1.

Result 5: $\dfrac{dS_{max}}{d\hat{\varepsilon}} = \mu$ **and that** $\dfrac{d\mu}{d\hat{\varepsilon}} < 0,$ **and** S_{max} **is a concave function of** $\hat{\varepsilon}$.

$$S_{max} = -\sum_{i=1}^{n} p_i \ln p_i$$

$$= -\sum_{i=1}^{n} p_i(-\lambda - \mu\varepsilon_i) \qquad \text{Using equation (1)}$$

$$= \lambda + \mu\varepsilon_i$$

$$\frac{dS_{max}}{d\hat{\varepsilon}} = \frac{d\lambda}{d\hat{\varepsilon}} + \mu + \hat{\varepsilon}.\frac{d\mu}{d\hat{\varepsilon}} \qquad\qquad (6)$$

Using equation (3), we get

$$e^\lambda = \sum_{i=1}^{n} e^{-\mu\varepsilon_i}$$

$$e^\lambda \frac{d\lambda}{d\hat{\varepsilon}} = \sum_{i=1}^{n} (-\varepsilon_i)e^{-\mu\varepsilon_i}\frac{d\mu}{d\hat{\varepsilon}}$$

$$= -\hat{\varepsilon}e^\lambda \frac{d\mu}{d\hat{\varepsilon}}$$

Using equation (6), we get

$$\frac{dS_{max}}{d\hat{\varepsilon}} = \mu$$

$$\frac{d^2 S_{max}}{d\hat{\varepsilon}^2} = (\sum_{i=1}^{n} e^{-\mu\varepsilon_i})^{-2}[(\sum_{i=1}^{n} e^{-\mu\varepsilon_i})(-\sum_{i=1}^{n}\varepsilon_i^2 e^{-\mu\varepsilon_i}) + (\sum_{i=1}^{n}\varepsilon_i e^{-\mu\varepsilon_i})^2]$$

$$= (\sum_{i=1}^{n} e^{-\mu\varepsilon_i})^{-2}[(\sum_{i=1}^{n}\varepsilon_i e^{-\mu\varepsilon_i})^2 - (\sum_{i=1}^{n} e^{-\mu\varepsilon_i})(\sum_{i=1}^{n}\varepsilon_i^2 e^{-\mu\varepsilon_i})] \le 0$$

So, S_{max} is a concave function of $\hat{\varepsilon}$.

4. APPLICATION OF MB DISTRIBUTION IN POLITICAL ENTROPY

Let $p_1, p_2, ..., p_n$ be the proportions of voters voting towards the political parties having attractive indices $a_1, a_2, ..., a_n$, respectively. Suppose that the only information about the system is that the average attractiveness is $\hat{a}$, i.e. we are given the information

$$\sum_{i=1}^{n} p_i a_i = \hat{a}$$

$$\sum_{i=1}^{n} p_i = 1, p_i \ge 0, \forall i = 1, 2, ..., n$$

Now, applying Jayne's Maximum Entropy principle, we choose that probability distribution that maximizes $-\sum_{i=1}^{n} p_i \ln p_i$, subject to the above constraints, the Lagrangian is

$$L \equiv -\sum_{i=1}^{n} p_i \ln p_i - (\lambda - 1)(\sum_{i=1}^{n} p_i - 1) - \mu(\sum_{i=1}^{n} p_i a_i - \hat{a})$$

Setting the derivative of the Lagrangian with respect to $p_1, p_2,..., p_n$ equal to zero, we get

$$-\ln p_i - p_i.\frac{1}{p_i} - (\lambda - 1) - \mu a_i = 0 \qquad\qquad i = 1, 2,..., n$$

$$\Rightarrow -\ln p_i - 1 - \lambda + 1 - \mu a_i = 0$$

$$\Rightarrow -\ln p_i = \lambda + \mu a_i$$

$$\Rightarrow \ln p_i = -\lambda - \mu a_i \qquad\qquad\qquad\qquad\qquad\qquad (7)$$

$$\Rightarrow p_i = e^{-\lambda - \mu a_i} \qquad\qquad\qquad\qquad\qquad\qquad\qquad (8)$$

Applying the natural constraint, we get

$$e^\lambda = \sum_{i=1}^n e^{-\mu a_i} \qquad\qquad\qquad\qquad\qquad\qquad (9)$$

Now from equation (7), we get

$$p_i = \frac{e^{-\mu a_i}}{\sum_{i=1}^n e^{-\mu a_i}} \qquad\qquad i = 1, 2,..., n$$

The distribution is the same as MB Distribution.
The Lagrange Multiplier μ is determined from

$$f(\mu) = [\frac{\sum_{i=1}^n a_i e^{-\mu a_i}}{\sum_{i=1}^n e^{-\mu a_i}}] - \hat{a} = 0 \qquad\qquad\qquad (10)$$

So that

$$\frac{d\hat{a}}{d\mu} = (\sum_{i=1}^n e^{-\mu a_i})^{-2}[(\sum_{i=1}^n e^{-\mu a_i})(-\sum_{i=1}^n a_i^2 e^{-\mu a_i}) + (\sum_{i=1}^n a_i e^{-\mu a_i})^2]$$

$$= (\sum_{i=1}^n e^{-\mu a_i})^{-2}[(\sum_{i=1}^n a_i e^{-\mu a_i})^2 - (\sum_{i=1}^n e^{-\mu a_i})(\sum_{i=1}^n a_i^2 e^{-\mu a_i})]$$

By using Cauchy – Schwarz's inequality, we find that $\frac{d\hat{a}}{d\mu} < 0$, unless $a_1 = a_2 = ... = a_n$ i.e. all the political parties have the same attractiveness, which is not the case. So, $\hat{a}$ is a monotonically decreasing function of μ .

Using the constraints, we get

$$d\hat{a} = \sum_{i=1}^n p_i da_i + \sum_{i=1}^n a_i dp_i \qquad\qquad\qquad\qquad (11)$$

Now,

$$S_{\max} = -\sum_{i=1}^{n} p_i \ln p_i$$

$$= -\sum_{i=1}^{n} p_i (-\mu a_i - \ln \sum_{i=1}^{n} e^{-\mu a_i}) \qquad (12)$$

$$= \mu \hat{a} + \ln \sum_{i=1}^{n} e^{-\mu a_i}$$

Political Temperature:

We take $\mu = \dfrac{1}{kT}$ where T may be defined as Political Temperature, and T is a monotonically increasing function of $\hat{a}$, i.e. average attractiveness. So, the greater the values of average attractiveness, the higher the political temperature is.

Case 1. If the political parties increase their attractiveness by intense publicity, but voters do not shift their loyalties:

In that case, dp_i becomes zero but da_i is increasing, so $d\hat{a}$ will increase, i.e., average attractiveness will increase, so political temperature will also increase.
From equation (10), we get

$$dS_{\max} = d\mu.\hat{a} + \mu.d\hat{a} - \frac{\sum_{i=1}^{n} a_i e^{-\mu a_i}}{\sum_{i=1}^{n} e^{-\mu a_i}} d\mu$$

$$= d\mu.\hat{a} + \mu.d\hat{a} - \hat{a}.d\mu$$

$$= \mu.d\hat{a}$$

So, in this case political temperature will rise but political heat will not be generated.

Case 2. If parties do not change their attractiveness, but voters shift their loyalties to more attractive parties:

Using equation (9), we get

$$dS_{\max} = \frac{1}{kT} (\sum_{i=1}^{n} p_i da_i + \sum_{i=1}^{n} a_i dp_i) \qquad (13)$$

Here da_i becomes zero.
From equation (11), we get

$$dS_{\max} = \frac{1}{kT} \sum_{i=1}^{n} a_i dp_i$$

$$= \frac{\Delta H}{kT}$$

In that case political temperature will rise and political heat is also generated.

REFERENCES

[1]   Huang, K., *Statistical Mechanics*, John Wiley & Sons, USA, 1928.

[2]   Clausius, R., Ueber eine veränderte Form des zweiten Hauptsatzes der mechanischen Wärmetheorie, in: *Abhandlungen Über Die Mechanische Wärmelehre*, Friedrich Vieweg und Sohn, Braunschweig, Germany, 1864.

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

[4]   Jaynes, E. T., "Information theory and statistical mechanics", *Physical Review,* 106 (1957) 620-630.

[5]   Kapur, J. N., *Maximum Entropy Models in Science and Engineering*, Wiley Eastern, New Delhi and John Wiley, New York, 1989.

[6]   Forte, B., and Sempi, C., "Maximizing conditional entropy: A derivation of quantal statistics", *Rendi Conta de Mathematics*, 9 (1976) 551-556.

[7]   Kapur, J.N., "Non-additive measures of entropy and distributions of statistical mechanics", *Ind. Jour. Pure App. Math*, 14 (11) (1977) 1372-1384.

[8]   Kapur, J.N., and Kesavan, H.K., "The generalized maximum entropy principle (with application)", Sandford Educational Press, University of Waterloo, Canada .

[9]   Kesavan, H.K., and Kapur, J.N., *Entropy Optimization Principle & their Application*, Academic Press, New York.

[10] Ray, A., and Sanat, K. M., "Derivation of some new distributions in statistical mechanics using maximum entropy approach", *Yugoslav Journal of Operations Research*, 24 (1) (2014) 27-34.

[11] Mazumder, S.K., "Maximum entropy and utility in a transportation system", *Yugoslav Journal of Operations Research*, 9 (1) (1999) 27-34.

[12] Ximing, W., "Calculation of maximum entropy densities with application to income distribution", *Journal of Econometrics*, 115 (2003) 347-354.