International Review of Pure and Applied Mathematics
Volume 15 • Number 1 • Jan-June 2019

A STUDY OF MAXIMUM ENTROPY PRINCIPLE UNDER WEIGHTED CONSTRAINT

Anita

Abstract: It is emphasized that the concept of weighted information has been proved to be very constructive because of its significance in goal oriented experiments. On the other hand, weighted mean has its own importance in the field of Statistics. The present communication is a step in the direction of the study of maximum entropy principle when weighted mean is prescribed. With the help of maximum entropy principle, we have made the study of optimizational problem related with generalized parametric measure of entropy.

Keywords: Probability distribution, Entropy, Uncertainty, Uniform distribution, Laplace's principle, Maximum entropy principle, Concave function.

1. INTRODUCTION

This is known fact in the literature information theoretic entropy that the uncertainty is maximum when the outcomes are equally likely. The uniform distribution maximizes the entropy, that is, the uniform distribution contains the largest amount of uncertainty, but this is just Laplace's principle of insufficient reasoning, according to which if there is no reason to discriminate between two or several events, the best strategy is to consider them as equally likely. Jaynes [4] gave a very natural criterion of choice by introducing the principle of maximum entropy. From the set of all probability distributions compatible with one or several mean values of one or several random variables, choose the one that maximizes Shannon's [14] entropy. Optimization includes maximization and minimization as well as simultaneous maximization of one function and minimization of another function.

A very natural question arises that why do we optimize entropy? We may make reduction in uncertainty by obtaining more and more information. Suppose we have a die and do not even know the number of faces it has. In this case, we may have a great deal of uncertainty and we may have many probability distributions

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

Further, we are given the information that the die has six faces. The uncertainty is reduced. We are now only limited to probability distribution , with (1.1) being satisfied. If, in addition, we are also given the mean number of points on the die, that is, we are given that

$$1p_1 + 2p_2 + 3p_3 + 4p_4 + 5p_5 + 6p_6 = 4.5 \quad (1.2)$$

Our choice of distribution is now restricted to those satisfying (1.1) and (1.2), and the uncertainty has been further reduced. If we are given the further information that

$$1^2 p_1 + 2^2 p_2 + 3^2 p_3 + 4^2 p_4 + 5^2 p_5 + 6^2 p_6 = 15, \quad (1.3)$$

then choice of distributions is further restricted and uncertainty is further reduced. We may go on getting more and more information and, accordingly, our uncertainty goes on decreasing. If we get in three stages three more independent linear constraints consistent with the equations (1.1), (1.2) and (1.3), we may get a unique set of values of p_1 through p_6. The uncertainty about these values is completely removed. At any stage, we may have infinity of probability distributions consistent with the given constraints, say,

$$\sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i g_{ri} = a_r, \ r = 1,2,...,m, \ m+1 < n \quad p_1 \geq 0,..., p_n \geq 0 \quad (1.4)$$

Out of these distributions, one has maximum uncertainty S_{max} and the uncertainty S of any other distribution is less than $S_{max.}$ Now uncertainty can be reduced by given additional information. Thus, the use of any distribution other than the maximum uncertainty distribution implies the use of some information in addition to that given by equation (1.4).

Kapur and Kesavan [9] has remarked that we should use all the information given to us and should avoid using any information not given to us. It is also the principle of scientific objectivity and honesty. According to the first part of the principle, we should use only distributions consistent with (1.4), but there may be infinity of such distributions. The second part of the principle now enables us to choose one out of these and we choose that which has the maximum uncertainty, S_{max}. Some other applications of measures of entropy for the study of maximum entropy principle have been provided by Kapur [8] whereas Parkash and Mukesh [10] have developed optimizational principle in the study of portfolio analysis.

It is to be noted that in the literature of information measures, there exist many probabilistic models of entropy or uncertainty. With the help of certain very plausible postulates which the uncertainty measure should possess, Shannon [14] investigated and characterized the first and foremost measure of entropy given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \quad (1.5)$$

The above result (1.5) holds good with the convention $0 \log 0 := 0$. Unless otherwise specified, it is to be noted that the base of logarithm is assumed to be 2.

Hence, for the first time, Renyi [13] presented the parametric group of entropies additive in nature as a mathematical generalization of Shannon's entropy, given by

$$H_\alpha(P)=\frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n}p_i^\alpha\right),\quad \alpha\neq 1,\alpha>0 \tag{1.6}$$

As $\alpha\to 1$, Renyi's entropy tends to Shannon's [14] entropy and is substantially more adaptable because of the parameter α, empowering several measurements of uncertainty within a given distribution. After Renyi's work, Havrda and Charvat [2] introduced the first non-additive entropy measure, given by

$$H^\alpha(P)=\frac{\left[\sum_{i=1}^{n}p_i^\alpha\right]-1}{2^{1-\alpha}-1},\alpha\neq 1,\alpha>0 \tag{1.7}$$

Tsallis [17] reinvented Havrda and Charvat's [2] entropy and specified it in the form

$$S_n^\alpha(P)=\frac{1}{1-\alpha}\left(\sum_{i=1}^{n}p_i^\alpha-1\right),\quad \alpha\neq 1,\ \alpha>0 \tag{1.8}$$

Some other well known measures of entropy investigated with the deep insight are explained below:

Kapur's [5, 6] entropies

$$_{\alpha,\beta}H(P)=\frac{1}{1-\alpha}\log\left(\frac{\sum_{i=1}^{n}p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n}p_i^{\beta}}\right),\quad \alpha\neq 1,\alpha>0,\beta>0,\ \alpha+\beta-1>0 \tag{1.9}$$

$$H_a(P)=-\sum_{i=1}^{n}p_i\ln p_i+\frac{1}{a}\sum_{i=1}^{n}\left[(1+ap_i)\ln(1+ap_i)-ap_i\right],\ a>0 \tag{1.10}$$

$$H_b(P)=-\sum_{i=1}^{n}p_i\ln p_i+\frac{1}{b}\left[\sum_{i=1}^{n}(1+bp_i)\ln(1+bp_i)+(1+b)\ln(1+b)\right],\ b>0 \tag{1.11}$$

The measure (1.9) reduces to Renyi's [13] measure when $\beta=1$, to Shannon's [14] measure when $\beta=1,\alpha\to 1$.

BURG'S [1] ENTROPY

$$S\left(P\right)=\sum_{i=1}^{n}\log p_i \tag{1.12}$$

SHARMA AND TANEJA'S [16] ENTROPIES

$$\Phi_1(P)=-2^{r-1}\sum_{i=1}^{n}p_i^r\log p_i\ ,\ r>0 \tag{1.13}$$

and

$$\Phi_2(P)=(2^{1-r}-2^{1-s})^{-1}\sum_{i=1}^{n}\left(p_i^r-p_i^s\right)\ ,\ r\neq s, r>0, s>0 \tag{1.14}$$

SHARMA AND MITTAL'S [15] ENTROPY

$$M^{\alpha}\left(P\right)=\frac{1}{2^{1-\alpha}-1}\left[\exp\left(\left(\alpha-1\right)\sum_{i=1}^{n}p_i\log p_i-1\right)\right]\ ,\ \alpha>0\ ,\alpha\neq 1 \tag{1.15}$$

Recently, Parkash and Mukesh [11, 12] developed the following entropy measures and provided their applications to Operation Research and Statistics:

$$H_1\left(P\right)=\frac{1}{\alpha}\sum_{i=1}^{n}\left(1-p_i^{\alpha p_i}\right),\ \alpha\neq 0,\alpha>0 \tag{1.16}$$

$$H_2\left(P\right)=\frac{1}{2\left(1-\alpha\right)}\sum_{i=1}^{n}\left[\alpha\left(p_i^{1-\alpha}-1\right)+\log p_i^{1-\alpha}\right],\alpha\neq 1,\ \alpha>0 \tag{1.17}$$

Many other measures of entropy have been discussed and investigated by Kapur [7], Herremoes [3], Sharma and Taneja [16], Parkash and Mukesh [11] etc.

In section 2, we have used Havrada-Charvat's [2] measure of order α and studied the principle of maximum entropy when weighted mean is prescribed.

2. MAXIMUM ENTROPY PRINCIPLE WHEN WEIGHTED MEAN IS PRESCRIBED

Let the random variable X takes the values $1,2,3,\ldots,n$ and the corresponding probabilities be $p_1,p_2,\ldots,p_n$. Let $\left(w_1,w_2,\ldots,w_n\right)$ be the weighted distribution. Then we want to find maximum entropy probability distribution when the weighted mean is prescribed to be m, where $1\leq m\leq n$. Such principle can be made applicable for any standard entropy and in the literature there are many such entropies already discussed above. But in the present problem, we make the study of Maximum Entropy Principle by using Havrada- Charvat's [1] entropy of second order. We find that when the weighted mean is prescribed as m, then for some value of m, the maximizing probabilities come out to be negative if we use the simple Lagrange's method of constrained optimization. This means that we have to take non-negativity constraints into account explicitly. Next, we do so and get the optimizing probability distribution.

Let a random variable takes values $1,\ 2,\ 3,...,n$. We want to find MEPD when Havrada Charvat's [2] measure of order 2 is maximized, subject to the weighted mean being prescribed as $m,\ 1 \le m \le n$.

Now Havrada-Charvat's [2] measure of order α is given by

$$_{\alpha}H\left(P\right)=\frac{1}{1-\alpha}\sum_{i=1}^{n}\left(p_i^{\alpha}-1\right) \tag{2.1}$$

For $\alpha = 2$, we have

$$_{2}H\left(P\right)=1-\sum_{i=1}^{n}p_i^2 \tag{2.2}$$

We want to maximize (2.2) subject to the constraints

$$\sum_{i=1}^{n}p_i=1,\ \ \sum_{i=1}^{n}w_i p_i=m,\ \ p_1,p_2,....,p_n\ge 0 \tag{2.3}$$

The corresponding Lagrangian is given by

$$L=1-\sum_{i=1}^{n}p_i^2+2\lambda\left(\sum_{i=1}^{n}p_i-1\right)+2\mu\left(\sum_{i=1}^{n}w_i p_i-m\right)$$

Now

$$\frac{\partial L}{\partial p_i}=-2p_i+2\lambda+2w_i\mu \tag{2.4}$$

For maxima, we equate (2.4) to zero and get

$$p_i=\lambda+w_i\mu;i=1,2,...,n \tag{2.5}$$

Equations (2.3) and (2.5) give

$$\sum_{i=1}^{n}(\lambda + w_i\mu) = 1 \text{ and } \sum_{i=1}^{n} w_i(\lambda + w_i\mu) = m \tag{2.6}$$

Under the weighted distribution $W = (w_1, w_2, w_3, ..., w_n) = (1, 2, 3, ..., n)$, equation (2.6) becomes

$$n\lambda + \frac{n(n+1)}{2}\mu = 1 \tag{2.7}$$

and

$$(\lambda + \mu) + 2(\lambda + 2\mu) + 3(\lambda + 3\mu) + ... + n(\lambda + n\mu) = m$$

that is,

$$\lambda(1 + 2 + 3 + ... + n) + \mu(1^2 + 2^2 + 3^2 + ... + n^2) = m$$

that is,

$$\lambda\frac{n(n+1)}{2} + \mu\frac{n(n+1)(2n+1)}{6} = m \tag{2.8}$$

From equation (2.4), we have

$$p_1 = \lambda + \mu \text{ and } p_n = \lambda + n\mu \tag{2.9}$$

where λ and μ can be calculated from (2.7) and (2.8).

Now from equation (2.7), we have

$$\lambda = \left(\frac{1}{n} - \frac{n+1}{2}\mu\right)$$

Using this value of λ in sequation (2.8), we get

$$\mu = \frac{12\left(m - \frac{n+1}{2}\right)}{n(n^2 - 1)}$$

This further gives $\lambda = \frac{1}{n}\left[1 - \frac{6\left(m - \frac{n+1}{2}\right)}{(n-1)}\right]$

Using these values of λ and μ in equation (2.9), we get

$p_1 = \lambda + \mu$

$$= \frac{12\left(m - \frac{n+1}{2}\right)}{n\left(n^2-1\right)} + \frac{1}{n}\left[1 - \frac{6\left(m - \frac{n+1}{2}\right)}{(n-1)}\right] = \frac{2}{n}\left[2 - \frac{3m}{n+1}\right]$$

Similarly, using the values of λ and μ , equation (2.9) gives

$$p_n = \frac{2}{n}\left[\frac{3m}{n+1} - 1\right]$$

Now these values of probabilities will be –ve if

$$\frac{2}{n}\left[2 - \frac{3m}{n+1}\right] < 0 \text{ or } \frac{2}{n}\left[\frac{3m}{n+1} - 1\right] < 0$$

that is, if $m < \frac{(n+1)}{3}$ or $m > \frac{2(n+1)}{3}$

Thus, Lagrange's method will be successful for this measure of entropy only when m has the following range:

$$\frac{n+1}{3} \le m \le \frac{2(n+1)}{3} \tag{2.10}$$

If $m < \frac{n+1}{3}$ or $m > \frac{2(n+1)}{3}$, this method fails.

Now if $m \le \frac{n+1}{3}$, we find that $p_n \le 0$ and so we set $p_n = 0$ and then choose $p_1, p_2, \ldots, p_{n-1}$ by using the principle of maximum entropy, we get

$$p_1 = \frac{2}{n-1}\left(2 - \frac{3m}{n}\right) \text{ and } p_{n-1} = \frac{2}{n-1}\left(\frac{3m}{n} - 1\right) \quad (2.11)$$

From the above representation, we conclude that:

(i) If $\frac{n}{3} < m \le \frac{n+1}{3}$, we get the MEPD in which only $p_n = 0$

(ii) If $\frac{n-1}{3} < m \le \frac{n+1}{3}$, we get the MEPD in which only p_{n-1} and p_n are zero.

(iii) If $\frac{n-2}{3} < m \le \frac{n+1}{3}$, we get the MEPD in which only p_{n-2}, p_{n-1}, p_n are zero and so on.

On the other hand, if $m > \frac{2(n+1)}{3}$, but $m < \frac{2(n+2)}{3}$, the MEPD will have only $p_1 = 0$.

If $\frac{2(n+1)}{3} < m < \frac{2(n+3)}{3}$, the MEPD will have only p_1 and p_2 equal to zero.

If $m = n$, then $p_1, p_2, ..., p_{n-1}$ are all zero.

NUMERICAL EXAMPLE: We illustrate the above method for $n = 8$. As discussed above, the method fails for $m < 3$ and $m > 6$.

We have $$\lambda = \frac{1}{n}\left[1 - \frac{6\left(m - \frac{n+1}{2}\right)}{n-1}\right] \text{ and } \mu = \frac{12\left(m - \frac{n+1}{2}\right)}{n\left(n^2 - 1\right)}$$

For $m = 3$, we have λ = 0.2857 and μ = -0.0357

Then $p_1 = \lambda + \mu = 0.2500$ $p_2 = \lambda + 2\mu = 0.2143$ $p_3 = \lambda + 3\mu = 0.1786$

$p_4 = \lambda + 4\mu = 0.1429$ $p_5 = \lambda + 5\mu = 0.1072$ $p_6 = \lambda + 6\mu = 0.0715$

$p_7 = \lambda + 7\mu = 0.0358$ $p_8 = \lambda + 8\mu = 0.000$

Obviously,

$$\sum_{i=1}^{8} p_i = 1.0003 \cong 1$$

Similarly, we have made the calculations of probabilities for $m = 3.5,\ 4.0,\ 4.5,\ 5.0,\ 5.5\ and\ 6.0$. The computations have been displayed in Table-2.1.

Table-2.1

m	P_1	p_2	P_3	p_4	P_5	p_6	p_7	P_8	$_2H_{max}$
3.0	0.2500	0.2143	0.1786	0.1429	0.1072	0.0715	0.0358	0.0000	0.8213
3.5	0.2083	0.1845	0.1607	0.1369	0.1131	0.0893	0.0655	0.0417	0.8512
4.0	0.1667	0.1548	0.1429	0.1310	0.1191	0.1072	0.0953	0.0834	0.8689
4.5	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.8750
5.0	0.0833	0.0952	0.1071	0.1190	0.1309	0.1428	0.1547	0.1666	0.8691
5.5	0.0142	0.0655	0.0893	0.1131	0.1369	0.1607	0.1845	0.2083	0.8527
6.0	0.0007	0.0364	0.0721	0.1078	0.1435	0.1792	0.2149	0.2506	0.8201

Thus we see that $_2H_{\max}(P)$ is a concave function and it assumes its maximum value at $m = 4.5$ as shown in Fig.-2.1.

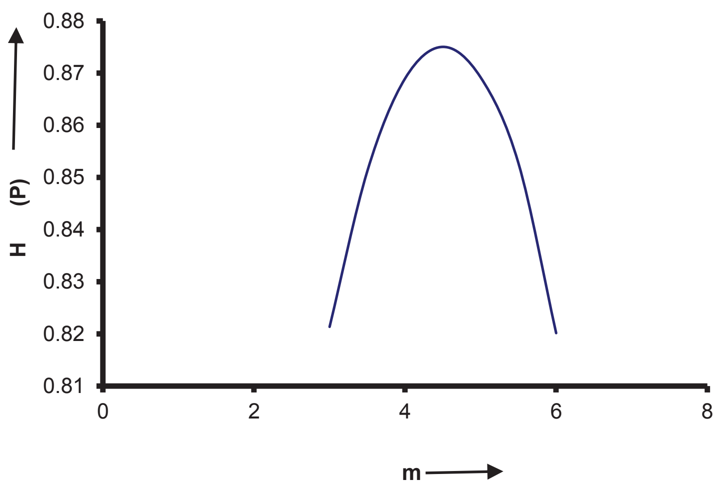

Fig.-2.1

Thus, from the above Fig.-2.1, we make the conclusion that when m = 4.5, the maximum entropy distribution is most uniform.

REFERENCES

[1] Burg, J. P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. In: Childrers, D. G. (ed). *Modern Spectral Analysis*. pp 130-131.

[2] Havrada, J.H. and Charvat, F. (1967). Quantification methods of classification process:Concept of structural α-entropy. *Kybernetika* 3: 30-35.

[3] Herremoes, P. (2006). Interpretations of Renyi entropies and divergences. *Journal of Physics A* 365: 57-62.

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

[5] Kapur, J.N. (1967). Generalized entropy of order α and type β. *Mathematics Seminar* 4: 79-84.

[6] Kapur, J.N. (1986). Four families of measures of entropy. *Indian Journal of Pure and Applied Mathematics* 17: 429-449.

[7] Kapur, J.N. (1995): "*Measures of Information and Their Applications*", Wiley Eastern, New York.

[8] Kapur, J.N. (1989). *Maximum Entropy Models in Science and Engineering*. Wiley Eastern, New Delhi.

[9] Kapur, J.N. and Kesavan, H.K. (1987). *The Generalized Maximum Entropy Principle*. Standard Educational Press.

[10] Parkash, O. and Mukesh (2012). Development of optimizational principle for minimizing risk in portfolio analysis. *Global and Stochastic Analysis: An International Journal* 2(1): 21-26.

[11] Parkash, O. and Mukesh (2012). New generalized parametric measure of entropy and cross entropy. *American Journal of Mathematics and Sciences* 1(1): 91-96.

[12] Parkash, O. and Mukesh (2015). Contribution of maximum entropy principle in the field of queueing theory. *Communications in Statistics-Theory and Methods* 45(12):3464-3472.

[13] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* 1: 547-561.

[14] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27: 379- 423, 623-659.

[15] Sharma, B. D. and Mittal, D. P. (1975). New non-additive measures of entropy for a discrete probability distributions. *Journal of Mathematical Sciences* 10: 122-133.

[16] Sharma, B. D. and Taneja, I. J. (1975). Entropies of type (α,β) and other generalized measures of information theory. *Metrica* 22: 202-215.

[17] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* 52: 479-487.

Anita
Department of Mathematics
Khalsa College For Women, Amritsar (INDIA)
Email: anitaarora@live.in