MEASURES OF UNCERTAINTY MATHEMATICAL PROGRAMMING AND PHYSICS§

J. N. KAPUR

*Indian Institute of Technology, Kanpur**

INTRODUCTION

The first section gives the measure of uncertainty given by Shannon (1948) and the generalizations thereof by Schützenberger (1954), Kullback (1959), Renyi (1951, 1965), Kapur (1967, 1968), and Rathie (1970). It gives some postulates characterizing Shannon's entropy, Renyi's entropy of order α and our entropy of order α and type β. It also gives some properties of this most general type of entropy.

In the second section an optimization problem is formulated and solved in the case of Shannon's and Renyi's entropies by the use of the principle of optimality. It is shown that this principle fails to solve the problem in the case of entropy of order α and type β and this leads to an interesting problem in non-linear integer fractional functional programming.

In the third section, we discuss the connection between the concepts of entropy in information theory and physics and show how Shannon's entropy leads to Boltzman distribution of statistical mechanics but fails to give the Fermi-Dirac and Bose-Einstein distributions of quantum mechanics. We find the entropies which lead to these distributions, but these do not satisfy an important property satisfied by Shannon's entropy. This may give us some insight into quantum mechanical systems.

In the fourth and last section, we obtain some properties of Bose-Einstein and Fermi-Dirac entropies obtained in the third section.

---

§ Contributed for the Silver Jubilee Session of the Indian Society of Agricultural Statistics.

* Present Address; Meerut University, Meerut.

2. Measures of Uncertainty

It is important to be able to measure the amount of information obtained from any scientific experiment or investigation. This can be measured by the amount of uncertainty removed by the experiment or investigation. The greater the uncertainty removed, the greater is the information communicated.

Thus let $A_1, A_2, \ldots, A_n$ be n possible outcomes of an experiment A and let $p_1, p_2, \ldots, p_n$ be the respective probabilities of these outcomes before the experiment is performed. After the experiment is performed, one of these outcomes has happened and so the uncertainty as to the outcome has been removed. We give below some postulates which we expect any measure of uncertainty to satisfy.

(*i*) It should be a function of $p_1, p_2, \ldots, p_n$. We denote it by $H(p_1, p_2, \ldots, p_n)$ or by $H_n(p_1, p_2, \ldots, p_n)$ if we want to indicate specifically the number of possible outcomes of the experiment.

(*ii*) If there are small changes in p's, there should result a small change in the measure of uncertainty, so that $H(p_1, p_2, \ldots, p_n)$ should be a continuous function of $p_1, p_2, \ldots, p_n$.

(*iii*) If $p_i = 1$ and the other probabilities are all zero, there is no uncertainty about the outcome. As such $H(p_1, p_2, \ldots, p_n)$ should vanish whenever one of the probabilities is unity and the others are zero.

(*iv*) From (*iii*), the minimum uncertainty is zero. The maximum uncertainty arises when all the outcomes are equally likely, *i.e.*, when

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n} \tag{1}$$

(*v*) If we add an impossible outcome to the n outcomes, the uncertainty does not change, so that

$$H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n). \tag{2}$$

(*vi*) If we simply interchange the names of the outcomes, the uncertainty does not change so that $H(p_1, p_2, \ldots, p_n)$ is a symmetric function of the arguments.

(*vii*) The information given by two independent experiments is the sum of informations given by the two separately so that if A and B are independent experiments, then

$$H(AB) = H(A) + H(B) \tag{3}$$

where

$$A \text{ is } \begin{pmatrix} A_1 A_2 \ldots A_n \\ p_1 p_2 \ldots p_n \end{pmatrix} \text{ and } B \text{ is } \begin{pmatrix} B_1 B_2 \ldots B_m \\ q_1\, q_2 \ldots q_m \end{pmatrix} \quad (4)$$

$$p_i \geqslant 0,\ q_j \geqslant 0\ (i=1, 2,\ldots, n; j = 1, 2,\ldots m) \quad (5)$$

and

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{j=1}^{m} q_j = 1. \quad (6)$$

(*viii*) It the experiments are not independent, we expect intuitively that

$$H(AB) \leq H(A) + H(B) \quad (7)$$

Let $H_k(B)$ be the amount of information given by experiment B when it is known that the experiment A has resulted in the kth outcome. We then postulate

$$H(AB) = H(A) + \sum_{k=1}^{n} p_k\, H_k(B). \quad (8)$$

It may be noted that here we have used a purely mathematical construct *viz* the concept of mathematical expectation.

We ask, at this stage, the question whether there exists a function satisfying all the eight postulates and whether it is unique. The answer to the first part is in the affirmative [Khinchin (1957)] and the function which satisfies all the postulates is

$$H(p_1, p_2,\ldots,p_n) = -\lambda \sum_{i=1}^{n} p_i \log p_i \quad (9)$$

where λ is arbitrary and this arbitrariness is due to the fact that none of our postulates specify a scale. If we add the postulate

(*ix*) $$H(\tfrac{1}{2},\tfrac{1}{2}) = 1, \quad (10)$$

we get a unique function *viz*

$$H(p_1, p_2,\ldots,p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i. \quad (11)$$

This is Shannon's entropy given in 1948.

We can prove a large number of properties for Shannon's entropy and we can replace the above postulate system by other postulate systems. Some of these postulate systems have been discussed in Reza (1961), Khinchin (1957), Feinstein (1958), Terberg (1958), Kullback (1959), Renyi (1961), Aczel (1968, 1970), Daroczy (1969), Kendell (1964), Lee (1964), and Aczel Daroczy (1971).

We may note that out of the above postulates, (*viii*) is on a different footing from others in the sense that it is less intuitive than others and is based on a mathematical construct. We can replace it by another postulate of the same nature, *e.g.* Renyi (1961) replaces it by

$$(viii') \qquad g[H_{n+m}(AUB)] = \frac{w(p)g(H_n(A)) + w(Q)g(H_m(B))}{w(P) + w(Q)} \tag{12}$$

where g is a monotone increasing function,

$$AUB \text{ is } \begin{Bmatrix} A_1A_2 \ldots A_n & B_1B_2 \ldots B_n \\ p_1p_2 \ldots p_n & q_1q_2 \ldots q_n \end{Bmatrix} \tag{13}$$

and

$$w(P) = p_1 + p_2 + \ldots + p_n \tag{14}$$

$$w(Q) = q_1 + q_2 + \ldots + q_m \tag{15}$$

are the weights of the two schemes.

Renyi (1961) considered the case of generalized probability distributions for which

$$w(P) \leq 1, \quad w(Q) \leq 1 \tag{16}$$

and he postulated (12) to hold when

$$w(P) + w(Q) \leq 1. \tag{17}$$

For these generalized probability distributions, he obtained the entropy of order α

$$H_\alpha(P) = \frac{1}{1-\alpha} \log_2 \frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i} \qquad (\alpha \neq 1). \tag{18}$$

We (1967) replaced the above postulate (*viii′*) by

$$(viii'') \qquad g(H_{n+m}(AUB)) = \frac{w_\beta (P) g(H_n (A)) + w_\beta (Q) g(H_m(B))}{w_\beta (P) + w_\beta (Q)} \qquad (19)$$

where

$$w_\beta (P) = p_1{}^\beta + p_2{}^\beta + \ldots + p_n{}^\beta \qquad \ldots(20)$$

$$w_\alpha (Q) = q_1{}^\beta + q_2{}^\beta + \ldots + q_m{}^\beta \qquad \ldots(21)$$

to get the entropy of order α and type β

$$H_\alpha^\beta (P) = \frac{1}{1-\alpha} \log_2 \frac{\sum_{i=1}^n p_i{}^{\alpha+\beta-1}}{\sum_{i=1}^n p_i{}^\beta} \quad (\alpha \neq 1). \qquad \ldots(22)$$

This will make the corresponding version of (17) applicable for a wider class of incomplete probability distributions.

We can get a still more general measure of uncertainty by using

$$w_f (P) = f(p_1) + f(p_2) + \ldots + f(p_n) \qquad (23)$$

$$w_f(Q) = f(q_1) + f(q_2) + \ldots + f(q_m) \qquad (24)$$

to get the measure

$$H^f_\alpha (P) = \frac{1}{1-\alpha} \log_2 \frac{\sum_{i=1}^n p_i{}^{\alpha-1} f(p_i)}{\sum_{i=1}^n f(p_i)} (\alpha \neq 1). \qquad \ldots(25)$$

This is the most general measure of uncertainty obtained so far.

If $f(p_i) = p_i{}^\beta$, we get entropy of order α and type β.

If $f(p_i) = p_i{}^\beta$, $\alpha \to 1$, we get the measure of uncertainty

$$H_1^\beta (P) = \underset{1 \to \alpha}{\text{Lt}} \; H_\alpha^\beta (P) = - \frac{\sum_{i=1}^n p_i{}^\beta \log_2 p_i}{\sum_{i=1}^n p_i{}^\beta}, \qquad \ldots(26)$$

If $f(p_i) = p_i$, we get Renyi's entropy of order α.

If $f(p_i) = p_i$, $\alpha \to 1$, $\sum_{i=1}^{n} p_i = 1$, we get Shannon's entropy.

If $\alpha \to \infty$, we get

$$H_{\infty}(P) = -\log_2 p_{max}. \tag{27}$$

which is in a sense, the simplest measure of uncertainty. It satisfies the first seven postulates and to use it we need not know all the probabilities; we have to know precisely only the maximum probability of any outcome. In many cases, it may be easier and less costly to determine this than to determine all the probabilities.

Rathie (1970) had generalized (22) to give

$$H_{\alpha}^{\beta_1 \beta_2, \dots \beta_n}(P) = \frac{1}{1-\alpha} \log_2 \frac{\sum_{i=1}^{n} p_i^{\alpha+\beta_i-1}}{\sum_{i=1}^{n} p_i^{\beta_i}} \qquad \dots(28)$$

but this measure violates the symmetry postulate (*vi*) which has a strong intuitive basis.

Before proceeding further, we state some properties of entropy of order α and type β, some of which will be found useful later.

(*i*) $H_{\alpha}^{\beta}(p_1, p_2, \dots, p_n)$ is indcpendent of both α and β and for generalized probability distributions with given weight k, it is a monotonic increasing function of n.

(*ii*) $$H_{\alpha}^{\beta}(P) = 0 \Rightarrow \sum_{i=1}^{n} p_i^{\beta}(p_i^{\alpha-1} - 1) = 0 \qquad \dots(29)$$

so that this entropy vanishes when exactly one probability is unity and the rest are all zero.

(*iii*) $H_{\alpha}^{\beta}(P)$ is a monotonic decreasing function of α, for a fixed β.

(*iv*) $$H_{\alpha}^{\beta}(p_1, p_2, \dots, p_n) \leq H_{\alpha}^{\beta}\left(\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}, p_3, \dots, p_n\right) \qquad (30)$$

$$H_{\alpha}^{\beta}(p_1, p_2, \dots, p_n) \leq H_{\alpha}^{\beta}\left(\frac{p_1+p_2+p_3}{3}, \frac{p_1+p_2+p_3}{3}, \frac{p_1+p_2+p_3}{3}, p_4, \dots, p_n\right) \qquad \dots(31)$$

and, in general, this measure increases with the coming closer of probabilities. This is also expected on intuitive grounds. Thus

$$\text{(v)}\quad H_{\alpha}^{\beta}\left(\frac{p_1}{m_1}, \frac{p_1}{m_1}, \ldots, \frac{p_1}{m_1}; \frac{p_2}{m_2}, \frac{p_2}{m_2}, \ldots, \frac{p_2}{m_2}; \ldots; \frac{p_n}{m_n}, \frac{p_n}{m_n}, \ldots, \frac{p_n}{m_n}\right) \geqslant H_{\alpha}^{\beta}(p_1, p_2, \ldots, p_n) \quad \ldots(32)$$

so that this entropy is always increased by a subdivision of the outcomes.

This last property leads to the optimization problem discussed below.

The entire theory of information is based on Shannon's entropy and Chabbra (1969) has made an investigation as to how far the results of this theory remain valid for entropy of order α and for entropy of order α and type β.

3. The Optimization Problem

Suppose an experiment

$$A=\begin{pmatrix} A_1, & A_2, \ldots, & A_n \\ p_1, & p_2, \ldots, & p_n \end{pmatrix}$$

is performed and $H(p_1, p_2, \ldots, p_n)$ is any one of the entropies obtained above, for this experiment.

Now suppose more funds become available so that each of the possible outcomes can be investigated in greater detail. The more detailed experiment is indicated by

$$\bar{A}\begin{pmatrix} A_{11}, A_{12}, \ldots, A_{1m_1} & A_{21}, A_{22}, \ldots, A_{2m_2} & A_{n1}, A_{n2}, \ldots, A_{nm_n} \\ p_{11}, p_{12}, \ldots, p_{1m_1} & p_{21}, p_{22}, \ldots, p_{2m_2} & p_{n1}, p_{n2}, \ldots, p_{nm_n} \end{pmatrix} \ldots(33)$$

where

$$\sum_{j=1}^{m_i} p_{ij}=p_i \quad (i=1, 2, \ldots, n). \qquad \ldots(34)$$

We have seen above that this will increase the information obtained. The problem is to maximize the gain in information *viz*

$$H(\bar{A})-H(A)$$

or

$$H(p_{11}, \ldots, p_{1m_1}\, ; p_{21}, \ldots, p_{2m_2}\, ; \ldots, p_{n1}, \ldots, p_{nm_n}) - H(p_1, p_2, \ldots, p_n) \qquad \ldots(35)$$

subject to

$$f_1(m_1) + f_2(m_2) + \ldots + f_n(m_n) \leq B \qquad \ldots(36)$$

where $f_i(m_i)$ is the cost of carrying out m_i experiments of the ith category, B is the budget available and m's are integers $\geqslant 1$.

Since m's are required to be integers, Lagrange's method of undetermined multipliers is not available, However fortunately the technique of dynamic programming is applicable in the case of Shannon's and Renyi's entropies. We may note that here decision or control variables are $m_1, m_2, \ldots, m_n$.

We may make an initial simplification by noting that the entropy is increased by making probabilities as nearly equal as possible so that we take the objective function as

$$H\left(\frac{p_1}{m_1}, \frac{p_1}{m_1}, \ldots, \frac{p_1}{m_1}\, ; \frac{p_2}{m_2}, \ldots, \frac{p_2}{m_2}; \ldots, \frac{p_n}{m_n}, \frac{p_n}{m_n}, \ldots, \frac{p_n}{m_n}\right) - H(p_1, p_2, \ldots, p_n). \qquad \ldots(37)$$

(i) In the case of Shannon's entropy, the objective function is

$$-\sum_{i=1}^{n} m_i \frac{p_i}{m_i} \log \frac{p_i}{m_i} + \sum_{i=1}^{n} p_i \log p_i = \sum_{i=1}^{n} p_i \log m_i. \qquad \ldots(38)$$

Let $\phi_n(B)$ denote the maximum of this objective function (38) subject to (36), m's being integers $\geqslant 1$, then the principle of optimality of dynamic programming [Bellman (1957), Bellman and Dreyfus (1961)] gives

$$\phi_n(B) = \max_{1 \leqslant m_n \leqslant M_n} [p_n \log m_n + \phi_{M-1}(B - f_n(m_n)]\ (n = 2, 3, \ldots) \qquad \ldots(39)$$

$$\phi_1(B) = p_1 \log M_1 \qquad \ldots(40)$$

where M_1 is the integer for which $B - f_1(M_1)$ is minimum and M_n is the integer for which

$$B - f_n(M_n) - f_1(1) - f_2(1), \ldots, -f_{n-1}(1)$$

is the smallest.

(*ii*) In the case of Renyi's entropy or order α, the objective function is

$$\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n} m_i\left(\frac{p_i}{m_i}\right)^{\alpha}}{\sum_{i=1}^{n} m_i\frac{p_i}{m_i}}-\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i} \quad \ldots(41)$$

so that if $0<\alpha<1$, we have to maximize $\sum_{i=1}^{n} p_i^{\alpha}/m_i^{\alpha-1}$ and if $\alpha>1$, we have to minimize the same quantity subject to (36) and m's being integers $\geqslant 1$.

The principle of optimality gives for the case $0<\alpha<1$

$$\phi_n(B)=\max_{1\leqslant m_n\leqslant M_n}\left[\frac{p_n^{\alpha}}{m_n^{\alpha-1}}+p_{n-1}(B-f_n(m_n)\right] \; (n=2,3,\ldots) \quad \ldots(42)$$

$$\phi_1(B)=\frac{p_1^{\alpha}}{M_1^{\alpha-1}} \quad \ldots(43)$$

Similar equations are obtained when $\alpha>1$.

(*iii*) In the case of $H_{\infty}(P)$, the objective function

$$-\log\max\left(\frac{p_1}{m_1},\frac{p_2}{m_2},\ldots,\frac{p_n}{m_n}\right)+\log\max(p_1,p_2,\ldots,p_n) \quad \ldots(44)$$

We have thus to find m_1, m_2, ..., m_n so that the largest of $\frac{p_1}{m_1}$, $\frac{p_2}{m_2}$, ..., $\frac{p_n}{m_n}$, is minimized so that the recurrence relation is

$$\phi_n(B)=\min_{1\leqslant m_n\leqslant M_n}\left\{\max\left(\frac{p_n}{m_n},\phi_{n-1}(B-f_n(m_n))\right)\right\} \; (n=2,3,..) \quad \ldots(45)$$

$$\phi_1(B)=\frac{p_1}{M_1}. \quad \ldots(46)$$

The particular cases of the above when (36) is replaced by

$$m_1+m_2+\ldots+m_n=M \quad \ldots(47)$$

were discussed by us in Kanpur (1968a).

(*iv*) For the entropy of order α and type β, the objective function is

$$\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n} m_i\left(\frac{p_i}{m_i}\right)^{\alpha+\beta-1}}{\sum_{i=1}^{n} m_i\left(\frac{p_i}{m_i}\right)^{\beta}}-\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n} p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} p_i^{\beta}} \quad \ldots(48)$$

so that we have to maximize (when $0<\alpha<1$)

$$\frac{\sum_{i=1}^{n} \frac{p_i^{\alpha+\beta-1}}{m_i^{\alpha+\beta-1}}}{\sum_{i=1}^{n} \frac{p_i^{\beta}}{m_i^{\beta-1}}} \quad \ldots(49)$$

subject to (36) and m's being integers $\geqslant 1$. When $\alpha>1$, we have to minimize the expression (49).

Unlike the earlier case, this is not separable and the technique of dynamic programming is not applicable.

The problem belongs to the domain of non-linear integer fractional functional programming.

The first attack is to try to solve this problem without the requirement of m's being integers. Even this is not easy and only some cases of fractional functional programming have been solved.

As is well-known, for a mathematical programming problem, whether a local maximum or minimum is also a global maximum or minimum, and the necessary and sufficient conditions for the existence of a local or global optimum depends very much on the nature of the objective and constraint-functions. Magasarian (1969) and Postein (1967) have studied seven types of convexity. Gupta and Bector (1967) have made a systematic study of the nature of products, quotients and rational powers of convex-like functions and their relevant results are given in the following table :

f	*CX* $\geqslant 0$	*SX* >0	*SX* >0	*CV* $\geqslant 0$	*CV* $\geqslant 0$	*SV* >0	*SV* >0	*CX* ≤ 0	*CX* ≤ 0	*SX* <0	*SX* <0
g	*SV* >0	*CV* >0	*SV* >0	*CX* >0	*SX* >0	*CX* >0	*SX* >0	*CX* >0	*SX* >0	*CX* >0	*CX* >0

$\frac{f}{g}$	EQX	EQX	EQX	QV	EQV	EQV	EQV	QX	EQX	EQX	EQX

f	CX $\geqslant 0$	CX $\geqslant 0$	SX >0	SX >0	CV $\geqslant 0$	CV $\geqslant 0$	SV >0	CV >0	CV ≤ 0	CV ≤ 0	SV <0	SV <0
g	CX <0	SX <0	CX <0	SX <0	CV <0	SV <0	CV <0	CV <0	CV >0	SV >0	CV >0	CV >0
$\frac{f}{g}$	QV	EQV	EQV	EQV	QX	EQX	EQX	EQV	QV	EQV	EQV	EQV

f	CX ≤ 0	CV ≤ 0	SX <0	CX <0	SX <0	SV <0	CV <0	SV <0
g	CV <0	CX <0	CV <0	SV <0	SV <0	CX <0	SX <0	SX <0
$\frac{f}{g}$	QV	QX	EQV	EQV	EQV	EQV	EQV	EQV

Here *CV*, *SV*, *QV*, *EQV*, *SDV* are abbreviations for concave, strictly concave, quasi-concave, explicitly quasi-concave and pseudo concave functions respectively. In the same way *CX*, *SX*, *QX*, *EQX* and *SDX* stand for convex, strictly convex, quasi-convex, explicitly quasi-convex and pseudo convex functions respectively. The results are simplified when the functions are differentiable.

4. Generalized Entropies and Statistical Mechanics

To make connection with statistical mechanics, we ask ourselves the question : "In what quantum state is the system ?" Let p_i be the probability for the system being in the ith state. The probability distribution must, of course, be consistent with the observed knowledge. According to the information-theoretic point of view, we choose a set of probabilities which maximize the entropy, consistent with observed knowledge, since this is the most non-committal view. According to Gibbs the state of equilibrium is the state of maximum entropy. From the information theoretic point of view, it is the state in which all the random motions which can take place are taking place so that the observer knows as little about the system as it is possible for him to know beyond the knowledge of the constants of motion.

From quantum mechanics, we know that the system can be in state i with energy ε_i ($i=1, 2, \ldots, n$). If we make an observation of ε, the best we can do so is to infer that it represents the expected energy $\bar{\varepsilon}$, since this is the repeatable quantity associated with the motion. To find the appropriate distribution, we maximize $H(p_1, p_2, \ldots\, p_n)$ subject to

$$\sum_{i=1}^{n} p_i = 1 \qquad \ldots(50)$$

$$\sum_{i=1}^{n} p_i \varepsilon_i = \bar{\epsilon} \qquad \ldots(51)$$

Equation (50) states that the system is always in one of the states and equation (51) states that $\bar{\varepsilon}$ is the expected energy.

Using Shannon's measure of uncertainty, we seek to maximize

$$-\sum_{i=1}^{n} p_i \log p_i \qquad \ldots(52)$$

subject to (50) and (51).

Using Lagrange's method of undetermined multipliers, we get the probability distribution

$$p_i = e^{-\psi - B_o \varepsilon_i} \quad (i=1, 2, \ldots, n) \qquad \ldots(53)$$

where ψ and B_o are determined from the equations

$$\sum_{i=1}^{n} e^{-B_o \varepsilon_i} = e^{\psi} \quad \sum_{i=1}^{n} \varepsilon_i e^{-B_o \varepsilon_i} = \bar{\varepsilon}\, e^{\psi}. \qquad \ldots(54)$$

It can be shown that B_o is equal to $1/kT$ where k is the Boltzman constant and T is the temperature of the system. (53) gives then the classical Boltzman distribution. However, in statistical mechanics, we come across the distributions

$$p_i = \frac{1}{e^{(\varepsilon_i - \mu)B_o} + 1} \quad \text{(Fermi-Dirac)} \qquad \ldots(55)$$

and

$$p_i = \frac{1}{e^{(\varepsilon_i - \mu)B_o} - 1} \quad \text{(Bose-Einstein)} \tag{56}$$

of which (54) is a limiting form when particle density is low and the temperature is high.

We now attempt to answer the following questions.

(*i*) How should the expression (52) for entropy be modified so that its maximization may lead to (55) or (56) instead of to (53)?

(*ii*) Do these modified forms satisfy the postulates laid down by Shannon for (52)? In particular which postulates have to be modified?

(*iii*) How are these modified forms related to the generalized entropies of order α and type β?

To answer these questions we replace (52) by

$$\sum_{i=1}^{n} f(p_i) \tag{57}$$

where the choice of function f is at our disposal. Maximizing it subject to (54), we get

$$f'(p_i) = B_o(\varepsilon_i - \mu). \tag{58}$$

If (55) is to be a solution, we get

$$f'\left(\frac{1}{e^{\mu}+1}\right) = \mu \text{ or } f'(y) = \log y\left(\frac{1}{y} - 1\right) = \log \frac{1-y}{y}$$

or

$$f(y) = -y \log y - (1-y) \log y + c. \tag{59}$$

The constant c depends on the choice of the base of the logarithm so that

$$S = -\sum_{i=1}^{n} p_i \log p_i - \sum_{i=1}^{n} (1-p_i) \log (1-p_i) \quad \ldots(60)$$

for the Fermi-Dirac case.

Similarly for the Bose-Einstein case, we get

$$\sum_{i=1}^{n} (1+p_i) \log (1+p_i) - \sum_{i=1}^{n} p_i \log p_i. \quad \ldots(61)$$

Let us now consider the properties of (60) and (61).

(*i*) Each is a continuous symmetric function of the arguments (taking $p_i \log p_i = 0$ when $p_i = 0$).

(*ii*) Each is maximum when $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$

[see Section 4].

(*iii*) For each $S(p_1, p_2, \ldots, p_n, o) = S(p_1, p_2, \ldots, p_n)$.

(*iv*) In the case of certainty, i.e, when one of the states is bound to occur and the rest are bound not to occur (60) gives zero and (61) gives 2 log 2, so we can make this also zero by defining

$$S = \sum_{i=1}^{n} (1+p_i) \log (1+p_i) - \sum_{i=1}^{n} p_i \log p_i - 2 \log 2. \quad \ldots (62)$$

(*v*) For two independent schemes, (60) gives

$$S(AB) - S(A) - S(B)$$

$$= -\sum_{j=1}^{m} \sum_{i=1}^{n} p_i q_j \log p_i q_j - \sum_{j=1}^{m} \sum_{i=1}^{n} (1 - p_i p_j) \log (1 - p_i q_j)$$

$$+ \sum_{i=1}^{n} p_i \log p_i + \sum_{j=1}^{m} q_j \log q_j + \sum_{i=1}^{n} (1-p_i) \log (1-p_j)$$

$$+ \sum_{j=1}^{m} (1-q_j) \log (1-q_j)$$

$$= -\sum_{j=1}^{m} \sum_{i=1}^{n} (1-p_i q_j) \log (1-p_i q_j) + \sum_{i=1}^{n} (1-p_i) \log (1-p_i)$$

$$+ \sum_{j=1}^{n} (1-q_j) \log (1-q_j). \quad \ldots (63)$$

In the special case

$$A = \begin{Bmatrix} A_1, & A_2, & \ldots, & A_n \\ \frac{1}{n} & \frac{1}{n} & & \frac{1}{n} \end{Bmatrix}, \quad B = \begin{Bmatrix} q_1, & q_2, & \ldots, & q_m \\ \frac{1}{m} & \frac{1}{m} & & \frac{1}{m} \end{Bmatrix}, \quad \ldots (64)$$

we get

$$
\begin{aligned}
S(AB)-S(A)-S(B) &= -nm\left(1-\frac{1}{nm}\right)\log\left(1-\frac{1}{nm}\right)\\
&\quad + n\left(1-\frac{1}{n}\right)\log\left(1-\frac{1}{n}\right)\\
&\quad + m\left(1-\frac{1}{m}\right)\log\left(1-\frac{1}{m}\right)\\
&= -(nm-1)\log\left(1-\frac{1}{nm}\right)\\
&\quad + (n-1)\log\left(1-\frac{1}{n}\right)\\
&\quad + (m-1)\log\left(1-\frac{1}{m}\right)\\
&= \log\frac{\left(1-\frac{1}{n}\right)^{n-1}\left(1-\frac{1}{m}\right)^{m-1}}{\left(1-\frac{1}{nm}\right)^{nm-1}}\\
&\neq 0 \qquad \ldots(65)
\end{aligned}
$$

so that we conclude that for two independent schemes $S(AB)$ is not necessarily equal to $S(A)+S(B)$. The same result is also easily seen to be true for Bose-Einstein entropy.

Thus we find that Fermi-Dirac and Bose-Einstein entropies do not satisfy postulates (*vii*) and (*viii*) of section 1.

However our generalized entropy of order α and type β and Renyi's entropy of order α satisfy (*vii*) and so it is observed that these cannot lead to (60) or (61).

The earlier generalizations referred to generalizations of postulate (*viii*). We need to modify postulate (*vii*) also.

For Renyi's entropy, it is easily seen that the probability distribution is given by

$$p_i=(A+B\,\varepsilon_i)^{\frac{1}{\alpha-1}} \qquad \ldots(66)$$

when A and B are determined from

$$\sum_{i=1}^{n}(A+B\,\varepsilon_i)^{\frac{1}{\alpha-1}}=1,\quad \sum_{i=1}^{n}\varepsilon_i\,(A+B\,\varepsilon_i)^{\frac{1}{\alpha-1}}=1. \qquad \ldots(67)$$

5. Bose-Einstein and Fermi-Dirac Entropies

(*i*) *We first prove by using the principle of optimality that Bose-Einstein entropy is maximum when all the probabilities are equal.* Let

$$S=\sum_{i=1}^{n}(c+p_i)\log(c+p_i)-\sum_{i=1}^{n}p_i\log p_i \qquad \text{...(68)}$$

and let it be desired to maximize this subject to

$$\sum_{i=1}^{n}p_i=c,\ p_i\geqslant 0. \qquad \text{...(69)}$$

Let $f_n(c)$ be the desired maximum value, then the principle of optimality gives

$$f_n(c)=\max_{0\leqslant p_n\leqslant c}\{(c+p_n)\log(c+p_n)-p_n\log p_n+f_{n-1}(c-p_n)\} \qquad (70)$$

$$f_1(c)=2c\log c-c\log c. \qquad \text{.. (71)}$$

Suppose the result is true for the case of $(n-1)$ outcomes, then (70) gives

$$f_n(c)=\max_{0\leqslant p_n\leqslant c}\Big\{(c+p_n)\log(c+p_n)-p_n\log p_n+(cn-p_n)\log\frac{cn-p_n}{n-1}-(c-p_n)\log\frac{c-p_n}{n-1}\Big\}$$

$$=\max_{0\leqslant p_n\leqslant c}\{\phi(p_n)\},\ \text{[say]} \qquad \text{...(72)}$$

then

$$\phi'(p_n)=\log\frac{c+p_n}{p_n}+\log\frac{c-p_n}{cn-p_n} \qquad \text{...(73)}$$

$$\phi''(p_n)=-\frac{c}{p_n(c+p_n)}-\frac{c(n-1)}{(c-p_n)(cn-p_n)}<0. \qquad \text{...(74)}$$

$\phi(p_n)$ is therefore maximum when

$$p_n=\frac{c}{n}. \qquad \text{...(75)}$$

Using the principle of induction and putting $c=1$, we get the result.

(*ii*) *We next prove, again by using the principle of optimality, that Fermi-Dirac entropy is maximized when all the probabilities are equal.*

We consider the maximization of

$$S=-\sum_{i=1}^{n} p_i \log p_i-\sum_{i=1}^{n} (c-p_i) \log (c-p_i) \quad \text{...(76)}$$

subject to (69)

The principle of optimality gives

$$f_n(c)=\max_{0\leqslant p_n\leqslant c} \{-p_n \log p_n-(c-p_n) \log (c-p_n)+f_{n-1}\ (c-p_n)\} \quad \text{...(77)}$$

$$f_1(c)=-c \log c \quad \text{...(78)}$$

If the result is true for the case of $(n-1)$ outcomes

$$f_n(c)=\max_{0\leqslant p_n<c} \Big\{-p_n \log p_n-(c-p_n) \log (c-p_n)-(c-p_n) \log \frac{c-p_n}{n-1}-(cn-2c+p_n) \log \frac{cn-2c+p_n}{n-1}\Big\}$$

$$=\max_{0\leqslant p_n\leqslant c} \{\phi(p_n)\} \qquad \text{(say).} \quad \text{...(78)}$$

Then

$$\phi'(p_n)=\log\frac{c-p_n}{p_n}+\log \frac{c-p_n}{cn-2c+p_n} \quad \text{...(79)}$$

$$\phi''(p_n)=-\frac{c}{p_n(c-p_n)}-\frac{cn-c}{(c-p_n)(cn-2c+p_n)}<0. \quad \text{...(80)}$$

$\phi(p_n)$ is therefore maximum when

$$p_n=\frac{c}{n}. \quad \text{...(81)}$$

Again using mathematical induction and putting $c=1$, we get the result.

(*iii*) *Alternative proofs of the above two results are easily obtained by using the convexity of the functions*

$$x \log x-(1+x) \log (1+x) \text{ and } x \log x+(1-x) \log (1-x) \quad \text{...(82)}$$

in the interval $0<x<1$ and the inequality

$$\phi\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)\leq \frac{1}{n}\sum_{i=1}^{n} \phi(a_i) \quad \text{...(83)}$$

which holds for any continuous convex function.

(*iv*) *We now show that both Bose-Einstein and Fermi-Dirac entropies are increased by subdivision of events.*

For the Bose-Einstein case, we have to show that

$$\left\{\sum_{i=1}^{n} m_i\left(1+\frac{p_i}{m_i}\right)\log\left(1+\frac{p_i}{m_i}\right)-\sum_{i=1}^{n} m_i\frac{p_i}{m_i}\log\frac{p_i}{m_i}\right.$$

$$\left.=\sum_{i=1}^{n}(1+p_i)\log(1+p_i)+\Sigma\, p_i\log p_i\right\}\geqslant 0$$

i.e.

$$\sum_{i=1}^{n}\left\{(p_i+m_i)\log\left(1+\frac{p_i}{m_i}\right)-(1+p_i)\log(1+p_i)+p_i\log m_i\right\}\geqslant 0. \quad ...(84)$$

Let

$$f(n)=(p+x)\log\left(1+\frac{1}{x}\right)-(1+p)\log(1+p)+p\log x \quad ...(85)$$

then

$$f'(x)=\log\left(1+\frac{p}{x}\right). \quad ...(86)$$

Since $f(1)=0$ and $f'(x)>0$ for $x\geqslant 1$, it follows that $f(x)\geqslant 0$ for $x\geqslant 1$. It follows that (84) is always satisfied when m's are integers $\geqslant 1$.

Similarly for the Fermi-Dirac case, we have to show that

$$\sum_{i=1}^{n}\left\{-(m_i-p_i)\log\left(1-\frac{p_i}{m_i}\right)+(1-p_i)\log(1-p_i)+p_i\log m_i\right\}\geqslant 0 \quad ...(87)$$

Let

$$\phi(x)=-(x-p)\log\left(1-\frac{p}{x}\right)+(1-p)\log(1-p)+p\log x \quad ...(88)$$

then

$$\phi'(x)=\log\frac{x}{x-p}. \quad ...(89)$$

Since $\phi(1)=0$, $\phi'(x)>0$ for $x\geqslant 1$, it follows that $\phi(x)\geqslant 0$ for $x\geqslant 1$. It follows then that (87) is always satisfied when m's are integers $\geqslant 1$.

(v) We now use dynamic programming to optimize the gain in Bose-Einstein and Fermi-Dirac entropies subject to the budget constraint (36).

The recurrence relations for the Bose-Einstein case are given by

$$\phi_n(B)=\max_{1\leqslant m_n\leqslant M_n} \{(p_n+m_n)\ \log\ \left(1+\frac{p_n}{m_n}\right)+p_n \log m_n+\phi_{n-1}(B-f_n(m_n))\}(n=2, 3,\ldots) \quad \ldots(90)$$

$$\phi_1(B)=(p_1+M_1)\log\left(1+\frac{p_1}{m_1}\right)+p_1 \log M_1. \quad \ldots(91)$$

The same relations for the Fermi-Dirac case are given by

$$\phi_n(B)=\max_{1\leqslant m_n\leqslant M_n} \{p_n \log m_n-(m_n-p_n)\log\left(1-\frac{p_n}{m_n}\right)+\phi_{n-1}(B-f_n(m_n))\} \quad ..(92)$$

$$\phi_1(B)=p_1 \log M_1-(M_1-p_1)\log\left(1-\frac{p_1}{M_1}\right). \quad \ldots(93)$$

Since the objective functions are separable, the method of dynamic programming is applicable.

6. Summary

The problem of allocating resources for carrying out experiments to maximize the gain in information when the budget is fixed is solved when Shannon's and Renyi's entropies are used.

For our entropy of order α and type β, this problem leads to a problem in non-linear integer fractional functional programming.

Shannon's entropy leads to Boltzman distribution of statistical mechanics. Bose-Einstein and Fermi-Dirac entropies which lead to the corresponding distributions of statistical mechanics have been obtained and some of their properties have been studied.

References

Aczel, J. (1968) : On Different Characterizations of Entropies, Proc. International Symp. Prob. and Inf. Theory, McMaster Univ., Canada, Lecture Notes in Math., Vol. 89, Spring, 1969, 1-11.

Aczel, J. (1970) : On Measures of Information and Their Characteristics, Proceedings of the Meeting on Information Measures, University of Waterloo, Canada.

Aczel, J. and Z. Daroczy (1971) : *Measures of Information and Their Characterizations*, Academic Press (to be published).

Bellman, R. (1959) : *Dynamic Programming*, Princeton University Press

Bellman, R. and S. E. Dreyfus (1961) : *Applied Dynamic Programming,* Princeton University Press.

Chabbra, P. (1969) : Contributions to Information Theory of Order α Ph.D. Thesis, I.I.T., Kanpur.

Daroczy, Z. (1969) : On the Shannon Measure of Information, Magyar Tud. Akad. Mat. Fiz. OSZT. Kozl., *19*, 9-24.

Fadeev, D.K. (1956) : On the Concept of Entropy of a Finite Probabilistic Scheme (Russian), Uspehi, Mat. Nauk *11,* 227-231.

Feinstein, A. (1958) : *Foundations of Information Theory,* McGraw-Hill, New York.

Gupta, S.K. and C.R. Bector (1967) : On the Nature of Products and Quotients of Functions, I.I.T., Kanpur Research Report.

Kapur, J. N. (1967) : Generalized Entropy of Order α and type β, The Math. Seminar *4,* 78-96.

Kapur, J.N. (1968*a*) : Some Applications of Dynamic Programming to Information Theory, Proc. Ind, Acad. Sc. *68A*, 1-11.

Kapur, J.N. (1968*b*) : Information of Order α and Type β, Proc. Ind. Acad. Sci. *68A*, 65-75.

Kapur, J.N. (1968*c*) : Some Properties of Entropy of Order α and Type β, Proc. Ind. Acad. Sci. *68A*, 201-211.

Kapur, J.N. (1968*d*) : Some Aspects of Mathematics of Operations Research, Presidential Address Mathematics Section, Indian Science Congress, 1-34.

Kendell, D.G. (1964) : Functional Equations in Information Theory, Z. Wahrscheinlichkeitstheorie und Veru. Gebiek *2,* 225-229.

Khinchin, A. (1957) : *Mathematical Foundations of Information Theory,* Dover, New York, 1-30.

Kullback, S. (1959) : *Information Theory and Statistics,* John Wiley, New York.

Lee, P.M. (1964) : On the Axioms of Information Theory, Ann. Math. Stat. *35*, 414-441.

Magasarian, O.L. (1964) : Pseudo Convex Functions, SIAM Control 3, 281-290.

Postern, J. (1967) : Seven Kinds of Convexity, SIAM Review *9,* 115-120.

Rathie, P.N. (1970) : On a Generalized Entropy and a Coding Theorem, J. Appl. Prob. 7.

Reza, F.M. (1961) : *Introduction to Information Theory*, McGraw-Hill.

Renyi, A. (1961) : On Measures of Entropy and Information, Proc. 4th Berkeley Symp. Math. Statist. and Prob., University of California Press, Berkeley *1,* 547-561.

Renyi, A. (1965) : On the Foundations of Information Theory, Rev. Ind. Internat. Statist. 33, 1-14.

Schutzenberger, M.P. (1954) Contributions aux Applications Statistiques de la theorie de l'information, Publ. Ind. Statist. Univ., Paris *3*, 3-117.

Shannon, C.E. (1968) : A Mathematical Theory of Communication, Bell System Tech. J. *27*, 379-423, 628-656.

Terberg, H. (1958) : A New Derivation of the Information Function, Math. Scand. *6*, 297-298.