Indian J. pure appl. Math., 17 (4) : 429-449 April 1986

FOUR FAMILIES OF MEASURES OF ENTROPY

J. N. KAPUR

Indian Institute of Technology, Kanpur 208016

(*Received* 7 *May* 1984; *after revision* 2 *August* 1985)

Four new families of measures of entropy are proposed. The concavity of each, both with respect to the probabilities $p_1, p_2, \cdots, p_n$ and with respect to the parameter involved is studied.

1. INTRODUCTION

Let $P = (p_1, p_2, \ldots, p_n)$ be a probability distribution. We are interested in finding a function $H_\alpha(p_1, p_2, \ldots, p_n)$, depending on a parameter α, with the following properties :

(i) It should be a non-negative continuous function of $p_1, p_2, \ldots, p_n$

(ii) It should be maximum when

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n}. \qquad \ldots(1)$$

(iii) The maximum value should increase with n

(iv) It should be minimum when $p_i = 1$, $p_j = 0$, $(j \neq i)$ for $i = 1, 2, \ldots, n$

(v) It should be a concave function of $p_1, p_2, \ldots, p_n$ so that when it is maximized, subject to linear constraints, its local maximum should be the same as its global maximum.

(vi) As the parameter α tends to zero, it should possibly approach Shannon's entropy $-\sum_{i=1}^{n} p_i \ln p_i$ or it should differ from the Shannon's entropy by a constant.

(vii) It should be either an increasing or a decreasing function of α.

(viii) It should be either a convex or a concave function of α.

We consider in the present paper, four functions with these properties.

We need such functions to measure the 'uncertainty' of a probability distribution or to measure the "information" provided by the realisation of an experiment, the probabilities of out-comes of which are given by this probability distribution.

The first condition is obviously essential since information should not be negative, it should change by a small amount when each probability changes by a small amount and it should not change when the outcomes are numbered differently.

The second and third conditions are required since for fixed n, the uncertainty should be maximum when all the outcomes are equally likely and this maximum uncertainty should increase with the number of outcomes.

The fourth condition is needed since when one of the outcomes is bound to occur, there is minimum uncertainty.

The fifth condition is needed since many of the applications of the entropy measure outside communication theory, arise through the application of the principle of maximum entropy[12,15,18]. In these applications, we maximize the measure of entropy subject to linear constraints, and if the measure is a concave function, we can be sure that a local maximum obtained with the help of Lagranges' method will also be the global maximum.

Since we are requiring our function to be both concave and symmetric, it will be Schur concave and as such if a distribution P is majorised by Q, then its entropy will be greater than that of Q (Marshall and Olkin[21]). This is again a highly desirable property for a measure of entropy or information.

The last three conditions are motivated by the only weakness of Shannon's[24] measure of entropy i. e. its inflexibility in use. Thus when it is maximized subject to the expected energy $\sum_{i=1}^{n} p_i \epsilon_i$, (where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are the energies of n states) being prescribed, we get Maxwell-Boltzmann distribution. But we have also Bose-Einstein and Fermi-Dirac distributions of statistical mechanics and we want to derive these also by using the principle of maximum entropy. In fact Kapur[14] got two special cases of the first measure proposed below, by a attempting to find the function $f(.)$ which is such that the measures $\sum_{i=1}^{n} f(p_i)$ when maximized subject to $\sum_{i=1}^{n} p_i \epsilon_i =$ constant gives either Bose-Einstein or Fermi-Dirac distribution. The first measure obtained below gives not only all these three distributions of statistical mechanics, but also a whole range of intermediate distributions[8,17], which can be used to discuss the transitional

behaviour. By maximizing other measures we get distributions which do not occur in Physics, but occur in economics, banking and sociology[18].

Similarly when Herniter[11] maximized Shannon's measure of entropy to estimate the probability of a customer changing from one brand of a product to another, Bass[2] criticized him because of the inflexibility of his results, since Herniters probability of transition was the same whether the product considered was soap or alchohal or whether the community belonged to Africa or USA. Later Kapur, Bector and Kumar[12] remedied this defect by using Renyi's[23] measure of entropy which depends on a parameter.

Most of the proposed measures of entropy[3,5,13-20-23,24,26] contain one, two or more parameters. These parameters give certain flexibility in applications and their values have ultimately to be determined from the data itself.

Another way of considering the need for parameters is to remember that the constraints of a physical problem proposed to be solved by the maximum entropy principle may not give all relevant information or the information may be such as may not be quantifiable in the form of constraints. Thus in the marketing problem, there are psychological, social and economic factors which in our present state of knowledge of human nature cannot be easily expressed as constraints. However the presence of two or more parameters can enable us to take at least partially, the variation due to these factors into account.

Still another reason for introducing a parameter is mathematical. Thus Campbell[6] defined a parametric exponential mean lengths of codewords so that he had one mean for each value of a parameter. The lower bounds for uniquely decipherable codes will also be in terms of an entropy depending on the parameter. In this case it happens to be Renyi's entropy.

Having established the need for parameters, we have to examine the conditions which these must satisfy. So far in the literature, the parameters have been obtained as the arbitrary constants in the solutions of certain functional equations. Here we have a little more freedom in the choice of the parameters and as such we can discuss the desirable conditions which can be required to be satisfied.

Condition (vi) arises from the consideration that in general Shannon's measure is a very desirable measure of information and as such as far as possible any family of measures of entropy should include. Shannon's measure as a limiting or as a particular case. Condition (vii) ensures that in general any data will be satisfied by only one value of the parameter. Condition (viii) is not so essential, but if it is satisfied, it will be very helpful if we want to ensure optimal fitting of data.

We find in the present paper four families of measures satisfying these eight highly desirable conditions. We do not claim that these are the only families of

measures satisfying these conditions. If there are more, the user will have a choice. He can impose additional conditions depending on his needs and he can choose the most appropriate measure for his problem. We feel that all the conditions should not be laid down from the purely mathematical context-free considerations and there should be some flexibility for the user to choose the measure which suits him most. This flexibility is provided by the occurrence of parameters as well as by the availability of a number of alternative measures of information.

Except Shannon's measures, there is almost no measure which is strictly unique in the sense that every measure depends on one or two parameters and the given axioms do not provide any guidelines for the determination of values of these parameters. Subject to this limitation, there is a uniqueness of a kind. However in many cases, axioms are chosen arbitrarily and no attempt is made to justify these. This applies particularly to generalised axioms of additivity. By using continuity and symmetry (which are highly desirable and necessary), recursivity (which is only a matter of manipulational convenience) and special forms of generalised additivity (for which justification is not easy to find) a unique measure is obtained. However quite often this measure is negative for some values of the parameters and for some probability distributions. Also in some cases the maximum entropy does not occur when all the probabilities are equal and the minimum value does not occur when one of the outcomes is certain. This is not surprising since these conditions are not included in the axiom system and cannot always be deduced from the given axiom system. In such cases, the measures obtained may be uniquely determined by the axioms, they may be nice mathematical objects, but they are almost useless as measures of information. Such measures have been severely criticized by Csiszer[7] and Aczel[1].

From the point of view of applications, uniqueness, for a set of arbitrarily stated set of axioms is not as important as the knowledge as to which properties a measure satisfies. It is equally important to know which properties it does not satisfy.

The only important property which our measures do not satisfy is that of additivity and thus these are non additive measures of entropy. These score over other non-additive measures because of the concave nature of these measures.

The two important additive measures are those of Shannon[24] and Renyi[23]. Our measures score over Shannon's measure since these contain a parameter and can give greater flexibility in applications. These score over Renyi's measure in the sense that unlike our measures, Renyi's measure when maximized subject to linear constraints, by the use of Lagrange's method can give rise to negative probabilities (Nathanson[22], Kapur[18]) which makes the use of Renyi's measure in some cases at least quite complicated as general methods of non-linear programming will have to be employed in such cases.

2. First Family of Entropy Measures

This is defined by

$$H_a(P) = H_a(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln p_i + \frac{1}{a}\sum_{i=1}^{n}[(1+ap_i) \times \ln(1+ap_i) - ap_i],\ a > 0 \quad \ldots(2)$$

(a) Let $f(x) = -x \ln x + \frac{1}{a}(1+ax)\ln(1+ax)$ ...(3)

so that

$$f''(x) = -[x(1+ax)]^{-1} < 0. \quad \ldots(4)$$

showing that $f(x)$ is a concave function of x. Using this fact and the result that the sum of any number of concave and linear functions is concave, we deduce that $H_a(P)$ is a concave function of $p_1, p_2, \ldots, p_n$.

(b) To maximize $H_a(P)$ subject to $\sum_{i=1}^{n} p_i = 1$, we use Lagrange's method to find that the maximum occurs when $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$ and the maximum value is given by

$$\ln n + \frac{1}{a}(n+a)\ln(n+a) - \frac{1}{a}(n+a)\ln n - 1 \quad \ldots(5)$$

Let

$$g(x) = \ln x + \frac{1}{a}(x+a)\ln(x+a) - \frac{1}{a}(x+a)\ln x - 1) \quad \ldots(6)$$

so that

$$g'(x) = \frac{1}{a}\ln\frac{x+a}{x} > 0 \quad \ldots(7)$$

showing that $g(x)$ is an increasing function of x. Using (5) we find that the maximum value of entropy is an increasing function of n.

(c) The minimum value of $H_\alpha(P)$ occurs at the vertices of the polyhedral region defined by

$$\sum_{i=1}^{n} p_i = 1,\ p_1 \geqslant 0,\ p_2 \geqslant 0,\ \ldots,\ p_n > 0 \quad \ldots(8)$$

i. e. at the n points $(1, 0, 0, \ldots, 0)$, $(0, 1, 0, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, 1)$. The value at each of these points is

$$\frac{1}{a}\,[(1+a)\ln(1+a)-a]. \qquad \ldots(9)$$

Let

$$h(x)=(1+x)\ln(1+x)-x \qquad \ldots(10)$$

so that

$$h(0)=0,\ h'(x)=\ln(1+x)>0,\ h(x)\geqslant 0. \qquad \ldots(11)$$

From (9) – (11), the minimum value of the entropy $\geqslant 0$ and as such $H_a(P)$ is always $\geqslant 0$. Also

$$\underset{a\to 0}{\text{Let}}\ \frac{(1+a)\ln(1+a)-a}{a}=\underset{a=0}{\text{Let}}\ \frac{\ln(1+a)}{1}=0 \qquad \ldots(12)$$

so that the minimum value of $H_a(P)$ is zero

$$\underset{a\to 0}{\text{Lt}}\sum_{i=0}^{n}\frac{(1+ap_i)\ln(1+ap_i)-a}{a}p_i=\underset{a\to 0}{\text{Lt}}\sum_{i=1}^{n}p_i\ln(1+ap_i)=0 \qquad \ldots(13)$$

so that

$$\text{(d)}\quad H_a(P)\to-\sum_{i=1}^{n}p_i\ln p_i \text{ as } a\to 0$$

$$\text{(e)}\quad \frac{d}{da}\,[H_a(P)]=\frac{1}{a^2}\sum_{i=1}^{n}[a\,p_i-\ln(1+a\,p_i)]. \qquad \ldots(14)$$

Let

$$h(x)=a\,x-\ln(1+ax) \qquad \ldots(15)$$

so that

$$h(0)=0,\ h'(x)=\frac{a^2x}{1+a\,x}>0,\ h(x)\geqslant 0. \qquad \ldots(16)$$

Thns $\frac{d}{da}\,[H_a(P)]>0$ when a >0 and as such $H_a(P)$ is a monotonic increasing function of a. Now

$$\left\{\frac{d}{da}\,[H_a(P)]\right\}_{a=0}=\underset{a\to 0}{\text{Lt}}\ \frac{\sum_{i=1}^{n}[a\,p_i-\ln(1+a\,p_i)]}{a^2}$$

(equation continued on p. 435)

$$= \operatorname*{Lt}_{a \to 0} \frac{\sum_{i=1}^{n} \left(p_i - \frac{p_i}{1 + a\, p_i} \right)}{2a}$$

$$= \operatorname*{Lt}_{a \to 0} \frac{\sum_{i=1}^{n} \frac{p_i^2}{(1 + a\, p_i)^2}}{2} = \frac{1}{2} \sum_{i=1}^{n} p_i^2 \,. \qquad \ldots(17)$$

Similarly

$$\left[\frac{d}{da} (H_a (P)) \right]_{a = \infty} = 0. \qquad \ldots(18)$$

As $a \to \infty$, $H_a (P) \to \infty$ and $\frac{d}{da}[H_a (P)] \to 0$

$$\text{(f)} \quad \frac{d^2}{da^2} [H_a (P)] = \frac{1}{a^3} \sum_{i=1}^{n} \left[2 \ln (1 + a\, p_i) - a\, p_i - \frac{a\, p_i}{1 + a\, p_i} \right]. \qquad \ldots(19)$$

Let

$$v (x) = 2 \ln (1 + x) - x - \frac{x}{1 + x'} \qquad \ldots(20)$$

so that

$$v (0) = 0, v' (x) = - \frac{x^2}{(1 + X)^2} < 0, v (x) \leqslant 0. \qquad \ldots(21)$$

From (19) —(21), $d^2/da^2 [H_a (P)] < 0$ and $H_a (P)$ is a concave function of a.

Thus $H_a (P)$ satisfies all the eight properties given in section 1. As a increases from 0 to ∞, $H_a (P)$ increases from $- \sum_{i=1}^{n} p_i \ln p_i$ to infinity, but it increases at a decreasing rate. Initially i. e. at $a = 0$, it increases at a rate $\frac{1}{2} \sum_{i=1}^{n} p_i^2$ and its final rate of increase is zero. Its graph is shown in Fig. 1.

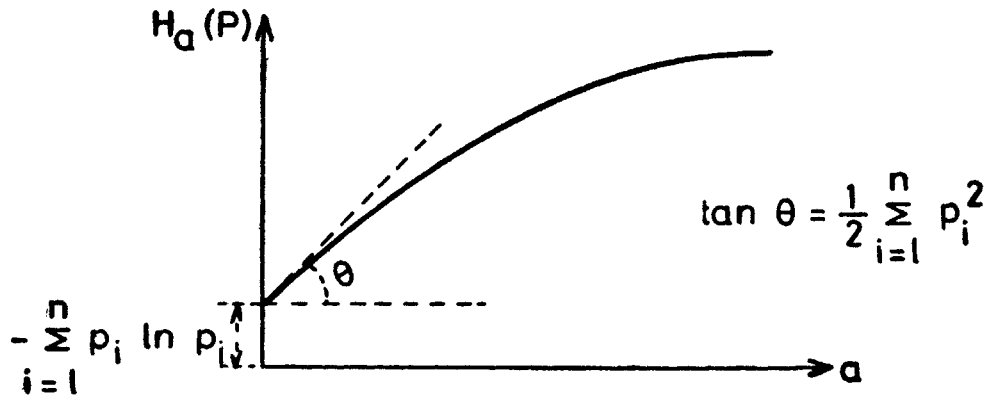

FIG. 1. Graph os $H_a(P)$ as a function of a.

3. SECOND FAMILY OF ENTROPY MEASURES

The minimum value of $H_a(P) > 0$. If we want the minimum value to be zero, we use the measure

$$H_b(P) = H_b(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln p_i$$

$$+ \frac{1}{b}\left[\sum_{i=i}^{n} (1 + bp_i)\ln(1 + bp_i) - (1 + b)\ \text{in}\ (1 + b)\right],$$

$$b \geqslant 0 \qquad \ldots(22)$$

$$\text{(a)}\ H_0(P) = -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{a\to 0} \frac{\sum_{i=1}^{n} (1+bp_i)\ \ln(1+bp_i) - (1+b)\ln(1+b)}{b}$$

$$= -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{b\to 0} \frac{\sum_{i=1}^{n} p_i \ln(1 + bp_i) - \ln(1 + b)}{1}$$

$$= -\sum_{i=1}^{n} p_i \ln p_i. \qquad \ldots(23)$$

$$H_\infty(P) = -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{b\to\infty} \sum_{i=1}^{n} p_i \ln \frac{1+bp_i}{1+b}$$

$$= -\sum_{i=1}^{n} p_i \ln p_i + \sum_{i=1}^{n} p_i \ln p_i = 0 \qquad \ldots(24)$$

$$\text{(b)}\quad \frac{d}{db}[H_b(p)] = \frac{1}{b^2}[\ln(1+b) - \sum_{i=1}^{n} (\ln(1+bp_i)]. \qquad \ldots(25)$$

Let

$$w(x) = \ln(1+bx) \qquad \ldots(26)$$

so that

$$w(0) = 0,\ W'(x) = \frac{b}{1+bx},\ W''(x) = -\frac{b^2}{(1+bx)^2} < 0. \qquad \ldots(27)$$

Thus $\ln(1+bx)$ is are concave function of x and $\sum_{i=1}^{n} \ln(1+bp_i)$ is a concave function of $p_1, p_2, \ldots, p_n$ of which the minimum value is $\ln(1+b)$ so that

$$\sum_{i=1}^{n} \ln(1+b\,p_i) > \ln(1+b). \qquad \ldots(28)$$

From (23), (24), (25), and (28), we find that $H_b(P)$ is a decreasing function of b which decreases from $-\sum_{i=1}^{n} p_i \ln p_i$ to 0 as b increases from 0 to ∞.

$$\text{(c)}\quad \frac{d^2}{db^2}[H_b(P)] = \frac{1}{b^3}\Big[2\sum_{i=1}^{n} \ln(1+b\,p_i) - \sum_{i=1}^{n} \frac{b\,p_i}{1+bp_i}$$

$$-2\ln(1+b) + \frac{b}{1+b}\Big]. \qquad \ldots(29)$$

Let

$$\alpha(x) = 2\ln(1+x) - \frac{x}{1+x'} \qquad \ldots(30)$$

so that

$$\alpha(0) = 0,\ \alpha'(x) = \frac{1+2x}{(1+x)^2} > 0,\ \alpha''(x) = \frac{-2x}{(1+x)^3} < 0. \qquad \ldots(31)$$

Thus $\alpha\ (x)$ is a concave function of x. Consequently

$$2 \sum_{i=1}^{n} \ln (1 + b\ p_i) - \sum_{i=1}^{n} \frac{b\ p_i}{1 + b\ p_i} \qquad \ldots(32)$$

is a concave function of $p_1, p_2, \ldots, p_n$ whose minimum value is

$$2 \ln (1 + b) - \frac{b}{1 + b} \ . \qquad \ldots(33)$$

Thus $d^2/\ db^2\ [H_b\ (P)] \geqslant 0$ and $H_b\ (P)$ is a convex function of $p_1, p_2 \ldots, p_n$.

From the above discussion and that of section 2, it follows that $H_b\ (P)$ satisfies all the conditions of section 1. Also

$$\left\{ \frac{d}{db}\ [H_b\ (P)] \right\}_{b=0} = \operatorname*{Lt}_{b \to 0} \frac{\ln (1 + b) - \sum_{i=1}^{n} \ln (1 + b\ p_i)}{b^2}$$

$$= \operatorname*{Lt}_{b \to 0} \frac{\frac{1}{1 + b} - \sum_{i=1}^{n} \frac{p_i}{1 + b\ p_i}}{2b}$$

$$= \operatorname*{Lt}_{b \to 0} \frac{-\frac{1}{(1 + b)^2} + \sum_{i=1}^{n} \frac{p_i^2}{(1 + b\ p_i)^2}}{2}$$

$$= \frac{-1 + \sum_{i=1}^{n} p_i^2}{2} < 0. \qquad \ldots(34)$$

Similarly

$$\left\{ \frac{d}{db} [H_b\ (P)] \right\}_{b=\infty} = 0. \qquad \ldots(35)$$

Thus $H_b\ (P)$ is a convex decreasing function of b which decreases from $-\sum_{i=1}^{n} p_i \ln p_i$ to 0 as b increases from 0 to ∞. Its initial rate of decrease is $\frac{1}{2}\ (1 - \sum_{i=1}^{n} p_i^2\ )$ and its final rate of decrease is zero. Its graph is shown in Fig. 2.

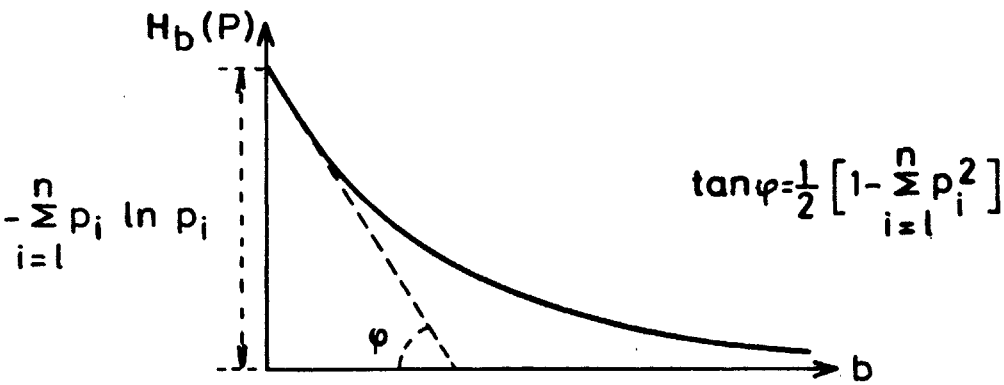

FIG. 2. Graph of $H_b(P)$ against b.

4. A THIRD FAMILY OF ENTROPY MEASURES

If we do not insist that $H_0(P)$ should differ from $-\sum_{i=1}^{n} p_i \ln p_i$ by a constant we can have other measure of entropy. We define

$$H_c(P) = -\sum_{i=1}^{n} p_i \ln p_i + \frac{1}{c^2}\left[\sum_{i=1}^{n}(1 + cp_i)\ln(1 + c\,p_i) - c\right] \quad \ldots(36)$$

$$\text{(a)}\quad H_0(P) = -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{c\to 0} \frac{\sum_{i=1}^{n}(1 + c\,p_i)\ln(1 + c\,p_i) - c}{c^2}$$

$$= -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{c\to 0} \frac{1}{2}\sum_{i=1}^{n} \frac{p_i^2}{1 + c\,p_i} = -\sum_{i=1}^{n} p_i \ln p_i$$

$$+ \frac{1}{2}\sum_{i=1}^{n} p_i^2 \quad \ldots(37)$$

Similarly

$$H_\infty(P) = -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{c\to\infty} \frac{1}{2}\sum_{i=1}^{n} \frac{p_i^2}{1 + c\,p_i} = -\sum_{i=1}^{n} p_i \ln p_i. \quad \ldots(38)$$

$$\text{(b)}\quad \frac{dH_c(P)}{dc} = \frac{1}{c^3}\sum_{i=1}^{n}[2cp_i - (2 + cp_i)\ln(1 + cp_i)]. \qquad \ldots(39)$$

Let

$$\beta(x) = 2x - (2 + x)\ln(1 + x) \qquad \ldots(40)$$

so that

$$\beta(0) = 0,\ \beta'(x) = \frac{x}{1 + x} - \ln(1 + x),\ \beta''(x) = -\frac{x}{(1 + x)^2} < 0. \qquad \ldots(41)$$

Thus $\beta'(x)$ is a decreasing function of x and since $\beta'(0) = 0$, $\beta'(x)$ is negative so that $\beta(x)$ is a decreasing function of x and again since $\beta(0) = 0$, $\beta(x)$ is negative and therefore from (39) $H_c(P)$ is a decreasing function of c. Again from (39)

$$\left\{\frac{d}{dx}[H_c(P)]\right\}_{c=0} = -\frac{1}{6}\sum_{i=1}^{n} p_i^3\ ,\ \left\{\frac{d}{dc}[H_c(P)]\right\}_{c=\infty} = 0 \qquad \ldots(42)$$

$$\text{(c)}\quad \frac{d^2}{dc^2}[H_c(P)] = \frac{1}{c^4}\sum_{i=1}^{n}[3(2 + cp_i)\ln(1 + cp_i) - 4cp_i - \frac{cp_i(2 + cp_i)}{1 + cp_i} - cp_i\ln(1 + cp_i)]. \qquad \ldots(43)$$

Let

$$\gamma(x) = 3(2 + x)\ln(1 + x) - 4x - \frac{x(2 + x)}{1 + x} - x\ln(1 + x)$$

$$= (6 + 2x)\ln(1 + x) - 5x - 1 + \frac{1}{1 + x} \qquad \ldots(44)$$

so that

$$\gamma(0) = 0,\ \gamma'(x) = \frac{-2x - 3x^2}{(1 + x)^2} + 2\ln(1 + x),\ \gamma'(0) = 0 \qquad \ldots(45)$$

$$\gamma''(x) = \frac{2x^2}{(1 + x)^3} > 0. \qquad \ldots(46)$$

$\gamma'(x)$ is an increasing function of x and $\gamma'(0) = 0$ so that $\gamma'(x) > 0$. Thus $\gamma(x)$ is an increasing function of x and $\gamma(0) = 0$ so that $\gamma(x) > 0$. Then from (43) $d^2/dc^2 [H_c(P)] \geqslant 0$ so that $H_c(P)$ is a convex function of c.

(d) It is easily shown that $-x \ln x + \frac{1}{c^2}(1 + cx) \ln (1+cx)$ is a concave function of x when $0 < x < 1$ and as such $H_c(P)$ is a concave function of $p_1, p_2, \ldots, p_n$.

(e) Thus $H_c(P)$ satisfies all the eight properties of section 1 except three $H_0(P)$ does not differ from $-\sum_{i=1}^{n} p_i \ln p_i$ by a constant. $H_\infty(P)$ differs from $-\sum_{i=1}^{n} p_i \ln p_i$ by a constant viz. zero.

$H_c(P)$ is a convex decreasing function of c which decreases from $-\sum_{i=1}^{n} p_i \ln p_i + \frac{1}{2}\sum_{i=1}^{n} p_i^2$ to $-\sum_{i=1}^{n} p_i \ln p_i$ as c increases from 0 to ∞. Its initial rate of decrease is $-\frac{1}{6}\sum_{i=1}^{n} p_i^3$ and its final rate of decrease is zero. Its graph is shown in Fig. 3.

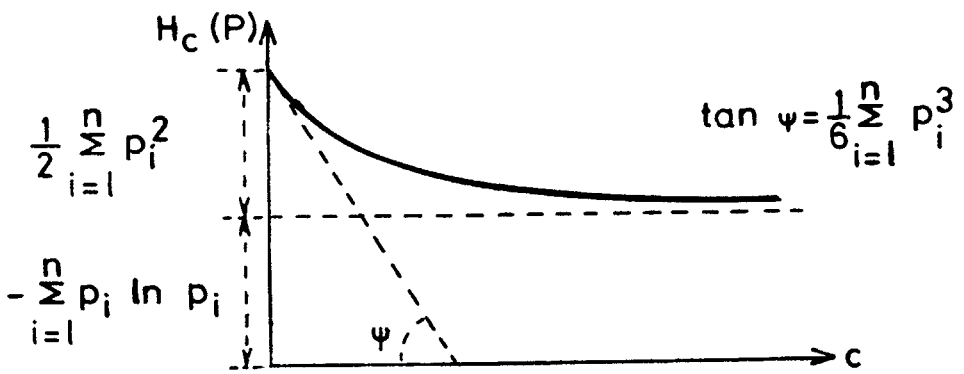

FIG. 3. Graph of $H_c(P)$ against c.

5. A FOURTH FAMILY OF ENTROPY MEASURES

We define it by

$$H_k(P) = -\sum_{i=1}^{n} p_i \ln p_i + \frac{1}{k^2}\left[\sum_{i=1}^{n} (1 + kp_i) \ln (1 + k p_i) - (1 + k) (\ln 1 + k)\right] \quad \ldots(47)$$

$$\text{(a)}\ H_0(P) = -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{k\to 0} \frac{\sum_{i=1}^{n}(1+kp_i)\ln(1+kp_i) - (1+k)\ln(1+k)}{k^2}$$

$$= -\sum_{i=1}^{n} p_i \ln p_i + \operatorname*{Lt}_{k\to 0} \frac{\sum_{i=1}^{n} \frac{p_i^2}{1+kp_i} - \frac{1}{1+k}}{2}$$

$$= -\sum_{i=1}^{n} p_i \ln p_i + \frac{\sum_{i=1}^{n} p_i^2 - 1}{2}. \qquad \ldots(48)$$

Similarly

$$H_\infty(P) = -\sum_{i=1}^{n} p_i \ln p_i \qquad \ldots(49)$$

$$\text{(b)}\ \frac{d}{dk}[H_k(P)] = \frac{1}{k^3}\sum_{i=1}^{n}(-2-kp_i)\ln(1+kp_i) + (2+k)\ln(1+k)]. \qquad \ldots(50)$$

Let

$$\delta(x) = (2+x)\ln(1+x) \qquad \ldots(51)$$

so that

$$\delta(0) = 0,\ \delta'(x) = \frac{2+x}{1+x} + \ln(1+x),\ \delta''(x) = \frac{x}{(1+x)^2} > 0. \qquad \ldots(52)$$

Thus $\delta(x)$ is a convex function of x, so that

$$\sum_{i=1}^{n}(2+kp_i)\ln(1+kp_i) \qquad \ldots(53)$$

is a convex function of $p_1, p_2, \ldots, p_n$ whose maximum value is

$$(2+k)\ln(1+k). \qquad \ldots(54)$$

Thus $d/dk\,[H_k(P)] \geqslant 0$ and $H_k(P)$ is an increasing function of k.

(c) To see whether $H_0(P) > 0$, we consider

$$\varphi(x) = -x \ln x - \frac{x - x^2}{2} \qquad \ldots(55)$$

so that

$$\varphi(0) = 0,\ \varphi'(x) = -3/2 + x - \ln x,\ \varphi''(x) = -1/x + 1 < 0,$$

$$\text{when } 0 < x < 1. \qquad \ldots(56)$$

Thus $\varphi(x)$ is concave, $\varphi'(x)$ is decreasing, but $\varphi'(1) = -\frac{1}{2} < 0$, so that $\varphi'(x) < 0$. Thus in $0 < x < 1$, $\varphi(x)$ is a decreasing function. Thus $H_0(P)$ is a concave function whose minimum value is zero so that $H_0(P) \geqslant 0$. Since $H_k(P)$ is an increasing function of k, $H_k(P)$ is also $\geqslant 0$.

$$\text{(d)}\ \frac{d^2H_k}{dk^2} = \frac{1}{k^2}\Big[6\sum_{i=1}^{n}(1 + k\,p_i)\ln(1 + k\,p_i) - 6(1+k)\ln(1+k)$$

$$- 4\sum_{i=1}^{n} k\,p_i \ln(1 + k\,p_i) + 4k\ln(1 + k)$$

$$+ \sum_{i=1}^{n} \frac{p_i^2\,k^2}{1 + kp_i} - \frac{k^2}{1 + k}\Big]. \qquad \ldots(57)$$

Let

$$\Psi(x) = 6(1 + x)\ln(1 + x) - 4x\ln(1 + x) + \frac{x^2}{1 + x} \qquad \ldots(58)$$

so that

$$\Psi(0) = 0,\ \Psi''(x) = \frac{2x^2}{(1 + x)^3} > 0. \qquad \ldots(59)$$

Thus $\Psi(x)$ is a convex function of x. As such

$$6\sum_{i=1}^{n}(1 + k\,p_i)\ln(1 + k\,p_i) - 4\sum_{i=1}^{n} k\,p_i \ln(1 + k\,p_i)$$

(equation continued on p. 444)

$$+\sum_{i=1}^{n}\frac{p_i^2\,k^2}{1+kp_i} \qquad \ldots(60)$$

is a convex function of $p_1, p_2, \ldots, p_n$ whose maximum value is

$$6(1+k)\ln(1+k) - 4k\ln(1+k) + \frac{k^2}{1+k} \qquad \ldots(61)$$

so that $d^2/dk^2\,(H_k) \leqslant 0$ and H_k is a concave function of k

$$\text{(e)}\quad \left(\frac{dH_k}{dk}\right)_{k=0} = \operatorname*{Lt}_{k\to 0}\frac{\sum_{i=1}^{n}(-2-kp_i)\ln(1+kp_i) + (2+k)\ln(1+k)}{k^3}$$

$$= \operatorname*{Lt}_{k\to 0}\frac{-2\sum_{i=1}^{n}\left(\frac{p_i}{1+k\,p_i}\right)^3 + \sum_{i=1}^{n}+\frac{p_i^3}{(1+k\,p_i)}\frac{2}{(1+k)^3} - \frac{1}{(1+k)^2}}{6}$$

$$= \frac{1-\sum_{i=1}^{n}p_i^3}{6}. \qquad \ldots(62)$$

Similarly

$$\left(\frac{d\,H_k}{dk}\right)_{k=\infty} = 0. \qquad \ldots(63)$$

The graph of $H_k\,(P)$ is shown in Fig. 4

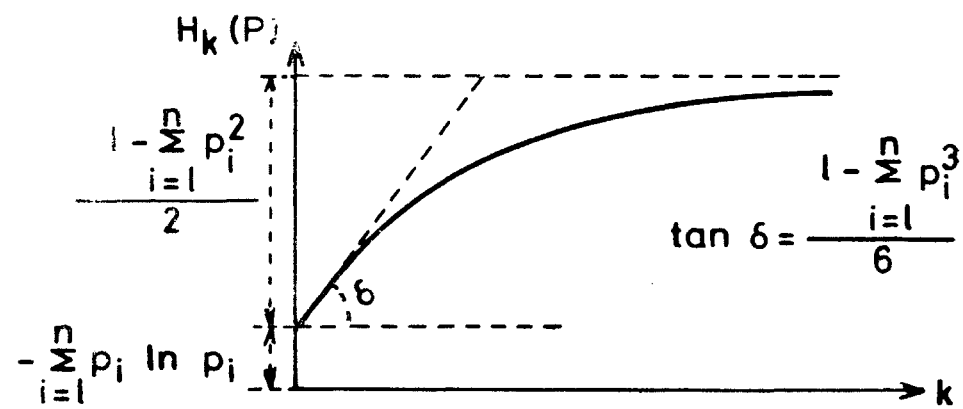

FIG 4. Graph of $H_k\,(P)$ against P.

Here also $H_0(P)$ differs from $-\sum_{i=1}^{n} p_i \ln p_i$ by a non-constant but $H_\infty(P)$ is same as $-\sum_{i=1}^{n} p_i \ln p_i$. All the other properties of section 1 are satisfied by $H_k(P)$.

6. Summary of Results

(i) Each of the four measures is a non-negative continuous function with minimum value at $(1, 0, \ldots, 0)$, $(0, 1, 0, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, 1)$ and maximum value at $(1/n, 1/n, \ldots, 1/n)$. The maximum value increases with n.

(ii) As a increase from 0 to ∞, $H_a(P)$ increases from $-\sum_{i=1}^{n} p_i \ln p_i$ to ∞ and it increases at a decreasing rate. Its initial rate of increase is $\frac{1}{2}\sum_{i=1}^{n} p_i^2$ and final rate of increase is zero.

(iii) As b increases from 0 to ∞; $H_b(P)$ decreases from $-\sum_{i=1}^{n} p_i \ln p_i$ to zero and it decreases at a decreasing rate. Its inital rate of decrease is $1 - \frac{1}{2}\sum_{i=1}^{n} p_i^2$ and final rate of decrease is zero.

(iv) As c increases from 0 to ∞, $H_b(P)$ decreases from

$$-\sum_{i=1}^{n} p_i \ln p_i + \frac{1}{2}\sum_{i=1}^{n} p_i^2 \text{ to } -\sum_{i=1}^{n} p_i \ln p_i .$$

It decreases at a decreasing rate. Its initial rate of decrease is $1/6 \sum_{i=1}^{n} p_i^3$ and final rate of decrease is zero.

(v) As k increases from 0 to ∞, $H_k(P)$, increases from

$$-\sum_{i=1}^{n} p_i \ln p_i - \frac{1 - \sum_{i=1}^{n} p_i^2}{2} \text{ to } -\sum_{i=1}^{n} p_i \ln p_i.$$

It increases at a decreasing rate. Its initial rate of increase is $\left(1 - \sum_{i=1}^{n} p_i^3\right)/6$ and final rate is zero.

(vi) Only $H_\alpha(P)$ is unbounded; all others are bounded. The final rate of change in each case is zero.

7. Comparison with Renyi's Measure of Entropy

Renyi's[1] measure of entropy is given by

$$H_\alpha(P) = \frac{1}{1-\alpha} \sum_{i=1}^{n} p_i^\alpha \quad , \alpha \neq 1, \alpha > 0. \qquad \ldots(64)$$

This is a non-negative continuous function of $p_1, p_2, \ldots, p_n$ which is strictly concave function of $p_1, p_2, \ldots, p_n$ for $0 < \alpha \leqslant 1$ and is pseudo-concave for all values of $\alpha > 1$. Its global maximum occurs when $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$ and the maximum value increases with n. Its minimum value is zero[5,13,18]. Further $H_\alpha(P)$ is a monotonic decreasing function[13] of α. Its maximum value occurs when $\alpha = 0$ and is given by $\ln n$ and its minimum value occurs when $\alpha \to \infty$ and is given by $-\ln p_M$ where p_m is max $(p_1, p_2, \ldots p_n)$. When $\alpha \to 1$, Renyi's entropy approaches Shannon's[24] entropy.

Our four measures have the advantage that these are concave functions of $p_1, p_2, \ldots, p_n$ for all $a, b, c, k > 0$.

Use of (64) in entropy maximizing models will only ensure that $\sum_{i=1}^{n} p_i^\alpha > 0$. It is possible that individual p_i's may be negative, when we use Lagrange's method, and $\sum_{i=1}^{n} p_i^\alpha$ may be positive. Thus Lagrange's method may lead to negative probabilities[18,22] and so the constraints $p_i \geqslant 0$ have to be explicitly used. In our case this situation cannot arise for our measures since they involve $\ln p_i$ and as such each p_i has to be > 0.

8. Extension of the Range of the Parameters for these Measures

We have considered only the ranges $a \geqslant 0, b \geqslant 0, c \geqslant 0, k \geqslant 0$ above but some of the properties obtained will continue to hold even for some negative values of the parameters. In general all the results will continue to hold for $a \geqslant -1$, $b \geqslant -1$, $c \geqslant -1$, $d \geqslant -1$.

Similarly we can extend Renyi's entropy for $\alpha < 0$. However we shall have here to restrict ourselves to distributions, all components of which are positive.

If $\alpha < 0$, x^α is convex, $\sum_{i=1}^{n} p_i^\alpha$ is convex, $\ln \left(\sum_{i=1}^{n} p_i^\alpha \right)$ is pseudo convex and

$$\frac{1}{\alpha - 1} \ln \left(\sum_{i=1}^{n} p_i^{\alpha} \right) \quad \ldots (65)$$

will be pseudo concave.

9. Use of our Measures to get Distributions of Statistical Mechanics

According to Jayne's[12] principle of maximum entropy, we maximize (2) subject to

$$\sum_{i=1}^{n} p_i, \sum_{i=1}^{n} p_i \epsilon_i = \bar{\epsilon} \quad \ldots (66)$$

to get

$$\ln \frac{1 + a p_i}{p_i} = \lambda + \mu \epsilon_i$$

or

$$\frac{1}{p_i} = e^{\lambda + \mu \epsilon_i} - a \text{ or } p_i = \frac{1}{e^{\lambda + \mu \epsilon_i} - a} \quad \ldots (67)$$

where λ and μ are determined by using (66). If $a = -1$, we get Fermi-Dirac distribution, if $a = 0$, we get Maxwell-Boltzamann distribution, if $a = 1$, we get Bose-Einstein distribution[8,15-18], other values of a between -1 and 1, we get Intermediate Statistics which are however different from Gentile[9] satistics. For $a > 1$, we get still newer statistics.

Maximizing (22), subject to (66) we get the same distributions

Maximizing (36) subject to (66), we get

$$-1 - \ln p_i + \frac{1}{c} (1 + \ln (1 + c p_i)) = \lambda + \mu \epsilon_i$$

or

$$\frac{1 + c p_i}{p_i^{\circ}} = e^{\lambda + \mu \epsilon_i} \quad \ldots (68)$$

which determines more complicated distributions.

10. Entropy Functions for $n = 2$

We consider the special case $n = 2, p_1 = p, p_2 = 1 - p$ so that

$$H_a(P) = -p \ln p - (1-p) \ln (1-p) + \frac{1}{a} [(1 + ap) \ln (1 + ap) + (1 + a - ap) \ln (1 + a - ap) - a]. \quad ...(69)$$

$$H_b(P) = -p \ln p - (1-p) \ln (1-p) + \frac{1}{b} [(1 + bp) \ln (1 + bp) + (1 + b - bp) \ln (1 + b - bp) - (1 + b) \ln (1 + b)] \quad ...(70)$$

$$H_c(P) = -p \ln p - (1-p) \ln (1-p) + \frac{1}{c^2} [(1 + cp) \ln (1 + cp) + (1 + c - cp) \ln (1 + c - cp) - c] \quad ...(71)$$

$$H_k(P) = -p \ln p - (1-p) \ln (1-p) + \frac{1}{k^2} [(1 + kp) \ln (1 + kp) + (1 + k - kp) \ln (1 + k - kp) - (1 + k) \ln (1 + k)]. \quad ...(72)$$

Each of these is of course symmetrical about $p = \frac{1}{2}$. The corresponding normalised measures of entropy i. e. measures which always lie between 0 and 1 are given by

$$\bar{H}_t(p) = \frac{H_t(p) - H_t(0)}{H_t(\frac{1}{2}) - H_t(0)}, \; t = a, b, c, k. \quad ...(73)$$

It is obvious that

$$\bar{H}_a(p) = \bar{H}_b(p), \; \bar{H}_c(p) = \bar{H}_k(p). \quad ...(74)$$

References

1. J. Aczel, *Inf. Proc. and Management*, **20** (1984) 383-93.
2. F. M. Bass, *J. Market Res.* **11** (1974), 1-20.
3. M. Behara, and J. S. Chawla Generalised gamma entropy : In *Selecta Statistica Candiana*, **2** (1984), 15-38.
4. M. Behara, and P. Nath, *Kybernetica* **10** (1974), 491-503.
5. Ben Basset and J. Raviv, *IEEE Trans. Inf. Th.*, It- **24** (1978), 324-31.

6. L. L. Campbell, *Inf. Control* **8** (1965), 423-25.
7. I. Csiszer, *Review of the book Measures of Information and their characterisation* by J. Aczel and Z. Daroczy," *IEEE Trans*, IT-22 (1976) 765-66.
8. B. Forte, and C. Sempi, *Rendi Contoide Mathematica* series VI **9** (1976), 551-66.
9. G. Gentile, *Nuovo Cim.*, **47** (1940) 493-97: **49** (1941), 106-10.
10. J. Havrda, and F. Charvat, *Kybernetica* **3** (1967) 30-35.
11. J. D Herniter, *J. Market Res.* **10** (1973) 361-75.
12. E. T. Jaynes *Phys. Rev.* **106** (1957), 620-30; **108**, 171-97.
13. J. N. Kapur, *Math. Sem.*, **4** (1967), 78-94.
14. J. N. Kapur, *J. Indian Soc. Ag. Stat.* **48A** (1972) 1-11.
15. J. N. Kapur, *J. Math. Phys. Sci.* **17** (1983), 103-56.
16. J. N. Kapur *Indian J. pure appl. Math.* **14** (1983), 1372-67.
17. J. N. Kapur, *Indian. J. pure appl. Math.* **14** (1983), 1433-38.
18. J. N. Kapur, *Maximum-Entropy Models in Science and Engineering*, South Asia Publishers, New Delhi, (1986).
19. J. N. Kapur, C. R. Bector, and U. Kumar, *Naval Res. Log. Quart.* **31** (1984), 183-98.
20. J. C. A. Vander Lubbe, Y-Bosima, and D. E. Boekee, *Infor. Sci.* **32** (1984), 187-215.
21. A. W. Marshall, and I. Olkin *Inequalities. Theory of Majorisatian and its Application.* Academic Press, New York, 1979.
22, M. Nathanson *Environ. Plann.* **10A** (1977), 879-88.
23. A. Renyi On Measures of entropy and information : In *Proc. Berkelay Symp. Math Stat Prob* **1** (1961), 547-60.
24. C. E. Shannon, *Bell Syst. Tech. J.* **27** (1948), 379-423, 623-59.
25. B. D. Sharma, and D. P. Mittal *J. Math. Soc.* **10** (1975), 38-40.
26. B. D. Sharma, and I. J. Taneja, *Electonishe Information Vonbert and Kybernatica* **13** (1977), 419-33.