Jñānābha, Vol. 17, 1987

A RESULT IN THE GEOMETRY OF PROBABILITY SPACE

By

J. N. KAPUR*

Department of Mathematics, Iadian Instiute of Technology,

Kanpur - 208016, India

(*Received* : *May 12, 1986*)

ABSTRACT

It ts shown that the set of those probability distributions p from which the weighted average of directed divergences to k given probability distributions $Q_1, Q_2, \ldots, Q_k$ is constant is also the set of distributions P from which the directed divergence to a fixed distribution Q is constant. In other words, it is shown that, in probability space, every 'ellipse' is a 'circle'.

1. BASIC DEFINITIONS

A *probability space* is the set of all probability distributions

$$P = (p_1, p_2, \ldots, p_n) \tag{1}$$

where

$$p_1 \geqslant 0, p_2 \geqslant 0, \ldots, p_n \geqslant 0, \sum_{i=1}^{n} p_i = 1. \tag{2}$$

Every probability distribution is represented by a point in ihis space. The point P is called a *non-degenerate point*, if none of the probabilities is zero; otherwise it is called a *degenerate point*.

***Current address** : Department of mathematics,

Institute of Technology, Hauz Khas, New Delhi-110016, India.

For different values of n, we get different probability spaces. A degenerate point in a higher dimensional space may be regarded as a non-degenerate point in a lower dimensional space. In our discussion, we shall consider a specified probability space *i.e,* we shall consider n to be fixed.

The *distance* of a point P from a point Q is measured by the Kullback-Leibler measure (see [8]) :

$$D (P : Q) = \sum_{i=1}^{n} p_i \ln (p_i/q_i) . \tag{3}$$

It can easily be shown [7] that $D(P:Q) \geqslant 0$ and $D(P:Q) = 0$ iff $P = Q$. It is also obvious that this distance is not symmetric since the distance of Q from P viz. $D(Q:P)$ is not necessarily equal to $D(P:Q)$, the distance of P from Q. It is also obvious that $D(P:Q)$ is finite only if either Q is non-degenerate or if Q is degenerate then P is also degenerate and corresponding to every zero component of Q, the corresponding component of P is also zero.

The posnts P corresponding to all those distributions P for which $D(P:Q) = K$ are said to lie on a *circle of the first type* with centre Q and radius K. In the same way the points corres-ponding to all those distributions P for which $D(Q:P) = K$ are said to lie on a *circle of the second type* with centre Q and radius K. These two circles are in general quite different.

The points P corresponding to all those distributions P for which $\lambda_1 D(P:Q_1) + \lambda_2 D(P:Q_2) + \ldots + \lambda_k D(P:Q_k) = K$, where Q_1, Q_2, . , Q_k are fixed probability distributions and $\lambda_1, \lambda_2, \ldots, \lambda_k$ are fixed non-negative constants whose sum is unity and K is a constant, are said to lie on a *weighted k-eilipse of the first iype* with focii at

$Q_1, Q_2, \ldots, Q_k$. It will be called a *weighted k-elliptic of the second type if* $\lambda_1 D(Q_1 : P) + \lambda_2 D(Q_2 : P) + \ldots + \lambda_k D(Q_k : P) = K$.

If $\lambda_1 = \lambda_2 = \ldots = \lambda_k$, these will be called *k-ellipses of the first and second type*. If $k = 2$, these will be called *ellipses of the first and second type*. If $k = 1$, these reduce to *circles of the first and second type*.

2. CIRCLES AND ELLIPSES OF THE FIRST TYPE

(a) *when Q is non-degenerate,* the circle of the first type is given by

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} = K. \tag{4}$$

Since for given non-zero $q_1, q_2, \ldots, q_n$, $\sum_{i=1}^{n} \ln(p_i/q_i)$ is a convex function of $p_1, p_2, \ldots, p_n$, its maximum value occurs at one of the vertices of the simplex (2) and

$$\max_{p} \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} = -\ln g_{min} \tag{5}$$

where

$$g_{min} = min(q_1, q_2, \ldots, q_n) \tag{6}$$

Thus if $k > -\ln g_{min}$, there is no probability distributions P satisfying (4) and as such the circle is imaginary. If $K = -\ln g_{min}$, there is just one probability distribution P satisfying (4), If $K = 0$, there is again one probability distribution viz. Q satisfying (4). When K lies between 0 and $-\ln g_{min}$, there may be an infinity of probability distributions satisfying (4) and as such there may be an infinity of points on the circle with centre Q and radius K.

(b) When $Q_1, Q_2, \ldots, Q_k$ are non-degenerate, the weighted

k-ellipses of the first type with focii at $Q_1, Q_2, \ldots, Q_k$ is given by

$$\lambda_1 D(P:Q_1) + \lambda_2 D(P:Q_2) + \ldots + \lambda_k D(P:Q_k) = K$$

or $$\sum_{j=1}^{k} \lambda_j \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_{ij}} = K \tag{7}$$

or $$\sum_{i=1}^{n} p_i \ln \frac{p_i}{\hat{q}_i} = K, \tag{8}$$

where

$$\hat{q}_i = \prod_{j=1}^{k} q_{ij}^{\lambda_j} \tag{9}$$

is the weighted geometrical mean of the i th components of $Q_1, Q_2, \ldots, Q_k$. Let

$$A = \sum_{i=1}^{n} \hat{q}_i . \tag{10}$$

From (8) and (10)

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{\hat{q}_i / A} = \sum_{i=1}^{k} p_i \ln A + K$$

or $$D(P:\tilde{Q}) = K' \tag{11}$$

where

$$\tilde{Q} = \left(\frac{\hat{q}_1}{A}, \frac{\hat{q}_2}{A}, \ldots, \frac{\hat{q}_n}{A} \right) \tag{12}$$

is a certain probability distribution whose i th component is proportional to the weighted geometrical mean of the ith components of $Q_1, Q_2, \ldots, Q_k$ and

$$K' = K + \ln A \tag{13}$$

is a new constant.

Now since the weighted geometric mean of k positive numbers $\leqslant$ their weighted arithmetic mean, we get from (9) and (10)

$$\hat{q}_i = \sum_{j=1}^{k} q_{ij}^{\lambda_j} \leqslant \sum_{j=1}^{k} \lambda_j q_{ij} ,$$

so that $A = \sum_{i=1}^{n} \hat{q}_i \leqslant \sum_{i=1}^{n} \sum_{i=1}^{k} \lambda_j q_{ij} = \sum_{j=1}^{k} \lambda_j \sum_{i=1}^{n} q_{ij} = \sum_{j=1}^{k} \lambda_j = 1,$ (14)

giving $A \leqslant 1$, $\ln A \leqslant 0$, $K' \leqslant K$. (15)

Now (11) represents a circle with centre $\tilde{Q}$ and radius K' . *Thus the weighted k–ellipse is a circle* . If the weights are $\lambda_1, \lambda_2, \ldots, \lambda_k$ and the focii of the k–ellipse are $Q_1, Q_2, \ldots, Q_k$, then the centre of the circle is $\tilde{Q}$ where $\tilde{Q}$ is defined by (9), (10), (12). The circle will be real only if

$$0 \leqslant K + \ln A \leqslant - \ln \frac{\min(\hat{q}, \hat{q}_2, \ldots, \hat{q}_n)}{A} . \tag{16}$$

If all the focii coincide, they will coincide with the centre of the circle.

If $K = -\ln A$, the radius of the circle is zero and thus $-\ln A$ gives the minimum value of $\lambda_1 D(P : Q_1) + \lambda_2 D(P : Q_2) + \ldots + \lambda_k D(P : Q_k)$ for variations of P so that

$$\min_{P} [\lambda_1 D(P : Q_1) + \lambda_2 D(P : Q_2) + \ldots + \lambda_k D(P : Q_k)] = -\ln A \tag{17}$$

Also $\max_{P} [\lambda_1 D(P : Q_1) + \lambda_2 D(P : Q_2) + \ldots + \lambda_k D(P : Q_k)]$

$$= \max_{P} [K] = \max_{P} [K' - \ln A]$$

$$= -\ln \frac{\min(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_n)}{A} - \ln A$$

$$= -\ln \min(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_n) \tag{18}$$

so that $-\ln \sum_{i=1}^{n} \hat{q}_i \leqslant \lambda_1 D(P:Q_1)+\lambda_2 D(P:Q_2)+\ldots+\lambda_k D(P:Q_k)$

$$\leqslant -\ln \min(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_n). \tag{19}$$

It may be noted that in Euclidean space, the weighted average of the distance if a point from k given fixed points has a minimum value, but no simple explicit expression is available for this and its maximum value is infinity .

(c) *When* Q *is degenerate* . If some components of Q are zero, then the corresponding components of P have also to be zero if $D(P:Q)$ is to be finite . If m components of Q are zero, then the same m components of P have to be zero and both P and Q belong essentially to $n-m$ dimensional probability space . The problem is thus reduced to a problem of type (a) in a lower dimensional space .

(d) *When some or all of* $Q_1, Q_2, \ldots, Q_k$ *are degenerate* . If the same components of $Q_1, Q_2, \ldots, Q_k$ are zero, the problem reduces to a problem in a lower dimensional space . If different components are zero, then all componennts of P corresponding to every zero components of $Q_1, Q_2, \ldots, Q_k$ have to be zero and P is highly restricted. In this case, it is better to use the second type of weighted ellipses .

3. CIRCLES AND ELLIPSES OF THE SECOND TYPE

(a) The distribution P belongs to the circle of the second type with centre Q and radius K if

$$\sum_{i=1}^{n} b_i \ln \frac{q_i}{p_i} = K . \tag{20}$$

If any component q_i = zero, then we need not take the corresponding component p_i to be zero, but if q_i is not zero, then p_i cannot be zero since k is finite .

(b) The weighted ellipse in this case is given by

$$\sum_{j=1}^{k} \lambda_j \sum_{i=1}^{n} q_{ij} \ln \frac{q_{ij}}{p_i} = K \tag{21}$$

$$\text{or } -\sum_{i=1}^{n} \left(\sum_{j=1}^{k} \lambda_j q_{ij} \right) \ln p_i = K - \sum_{j=1}^{k} \sum_{i=1}^{n} \lambda_j q_{ij} \ln q_{ij}$$

$$\text{or } \quad -\sum_{i=1}^{n} \bar{q}_i \ln p_i = K - \sum_{j=1}^{k} \sum_{i=1}^{n} \lambda_j q_{ij} \ln q_{ij} . \tag{22}$$

Tnis can be compared with the equation of the circle

$$-\sum_{i=1}^{n} q_i \ln p_i = K - \sum_{i=1}^{n} q_i \ln q_i . \tag{23}$$

$$\text{Now } \sum_{i=1}^{n} \bar{q}_i = \sum_{i=1}^{n} \sum_{j=1}^{k} \lambda_j q_{ij} = \sum_{j=1}^{k} \lambda_j \sum_{i=1}^{n} q_{ij} = \sum_{j=1}^{k} \lambda_j = 1, \tag{24}$$

$$\text{so that } \quad \bar{Q} = (\bar{q}_1 , \bar{q}_2 , \ldots , \bar{q}_n) \tag{25}$$

is a proper probability distribution . In fact

$$\bar{Q} = \lambda_1 Q_1 + \lambda_2 Q_2 + \ldots + \lambda_k Q_k , \tag{26}$$

is the weighted arithmetic mean of $Q_1 , Q_2 , \ldots , Q_k$. From (22) and (23) we find that the weighted k-ellipse of the second type with focii $Q_1 , Q_2 , \ldots , Q_k$ is a circle of the second type with centre $\bar{Q}$. Its radius is given by

$$R = \sum_{i=1}^{n} \bar{q}_i \ln \bar{q}_i + k - \sum_{j=1}^{k} \sum_{i=1}^{n} \lambda_j q_{ij} \ln q_{ij} \tag{27}$$

$$= - H(\bar{Q}) + K + \sum_{j=1}^{k} \lambda_j H(Q_j) , \tag{28}$$

where $H(P) = - \sum_{i=1}^{n} p_i \ln p_i$ (29)

is Shannon's [10] measure of entropy so that

$$R = K - [H(\lambda_1 Q_1 + \lambda_2 Q_2 + \dots + \lambda_k Q_k) - \lambda_1 H(Q_1) - \dots - \lambda_k H(Q_k)] \tag{30}$$

$$= K - J(\lambda_1, \lambda_2, \dots, \lambda_k ; Q_1, Q_2, \dots, Q_k) , \tag{31}$$

where $J(\lambda_1, \lambda_2, \dots, \lambda_k ; Q_1, Q_2, \dots, Q_k)$ is Jensen's difference [1] . Since $H(P)$ is a concave function of $p_1, p_2, \dots, p_n$, this Jensen's difference $\geqslant 0$ and vanishes iff $Q_1 = Q_2 = \dots = Q_k$.

Thus the weighted ellipse is a circle whose centre is the weighted arithmetic mean of $Q_1, Q_2, \dots, Q_k$ and whose radius $\leqslant K$ and is equal to K iff $Q_1 = Q_2 = \dots = Q_k$ *i.e.* iff all the focii coincide. The circle will be real iff

$$K \geqslant J(\lambda_1, \lambda_2, \dots, \lambda_k , Q_1, Q_2, \dots, Q_k) \tag{32}$$

If $K = J(\lambda_1, \lambda_2, \dots, \lambda_k ; Q_1, Q_2, \dots, Q_k)$, the circle become a point circle consisting of only one point viz $\bar{Q}$.

4. COMPARISON OF THE TWO TYPES OF WEIGHTED ELLIPSES

For both types, the weighted ellipse is a circle. The centres in the two cases are given by

$$\tilde{Q} = \left(\frac{\hat{q}_1}{A}, \frac{\hat{q}_2}{A}, \dots, \frac{\hat{q}_n}{A} \right) \tag{33}$$

$$\tilde{Q} = (\bar{q}_1, \bar{q}_2, \ldots, \bar{q}_n) \qquad (34)$$

The first centre is based on the weighted geometrical means of the i th components of Q_1, Q_2, ..., Q_k and the second centre is based in their weighted arithmetic means. The radii of the two circles are

$$R_1 = K + \ln \left(\sum_{i=1}^{n} \hat{q}_i \right) \qquad (35)$$

and $$R_2 = K - [H (\lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_k P_k) - \sum_{j=1}^{k} \lambda_j H (P_j)]. \qquad (36)$$

The radii in both cases are less than K. In the first case the difference is the logarithm of the sum of the geometrical means and in the second case it is the Jensen's difference.

In the first case, there is an upper bound on K, in the second case there is none.

5. CONTINUOUS-VARIATE DISTRIBUTIONS CASE

(a) Let P, Q, Q_1, Q_2, .., Q_k distributions have density functions $f(x)$, $g(x)$ $g_1(x)$, $g_2(x)$, ..., $g_k(x)$ respectively, then the first type of weighted ellipse is given by

$$\sum_{j=1}^{k} \lambda_j \int f(x) \ln \frac{f(x)}{g_j(x)} \, dx = K \qquad (37)$$

or $$\int f(x) \ln \frac{f(x)}{\hat{g}(x)} \, dx = K ; \quad \hat{g}(x) = \prod_{j=1}^{k} g_j^{\lambda_j}(x) \qquad (38)$$

or $$\int f(x) \ln \frac{f(x)}{\hat{g}(x)/A} = K + \ln A ; \; A = \int \hat{g}(x) \, dx \qquad (39)$$

which is circle with centre of probability distribution with density function $\hat{g}(x)/A$ and with radius $K + \ln A$.

Since the geometric mean of a number of positive numbers $\leqslant$ the arithmetic mean

$$\hat{g}(x) \leqslant \sum_{j=1}^{k} \lambda_j g_j(x), \tag{40}$$

so that

$$A \leqslant \int \sum_{j=1}^{k} \lambda_j g_j(x)\, dx = \sum_{j=1}^{k} \lambda_j \int g_j(x)\, dx = \sum_{j=1}^{k} \lambda_j = 1 \tag{41}$$

and $$\ln A \leqslant 0. \tag{42}$$

(b) The weighted ellipse of the second type is given by

$$\sum_{j=1}^{k} \lambda_j \int g_j(x) \ln \frac{g_j(x)}{f(x)}\, dx = K \tag{43}$$

or

$$-\int \bar{g}(x) \ln f(x)\, dx \int K - \sum_{j=1}^{k} \lambda_j \int g_j(x) \ln g_j(x)\, dx \tag{44}$$

where

$$\bar{g}(x) = \sum_{j=1}^{k} \lambda_j g_j(x), \tag{45}$$

so that

$$\int \bar{g}(x) \ln \frac{\bar{g}(x)}{f(x)}\, dx = \int \bar{g}(x) \ln \bar{g}(x)\, dx + K - \sum_{j=1}^{k} \lambda_j \int g_j(x) \ln g_j(x)\, dx$$

$$= K - (H(\bar{g}(x)) - \sum_{j=1}^{k} \lambda_j H(g_j(x)) \tag{46}$$

$$= K - J(\lambda_1, \lambda_2, \ldots, \lambda_k, g_1(x), g_2(x), \ldots, g_k(x)) \tag{47}$$

which is a circle of the second type with centre at the probability distribution with density function $\bar{g}(x)$ and with radius less than K by an amount equal to Jensen's difference .

These results are similar to these for the discrete-variate distributions ,

6. USE OF OTHER MEASURES OF DIRECTED DIVERGENCE

(a) If we use Havrda and Charvat's [2] measure of directed divergence, the weighted ellipse of the first type is

$$\frac{1}{\alpha-1} \sum_{j=1}^{k} \lambda_j \left(\sum_{i=1}^{n} p_i^{\alpha} q_{ij}^{1-\alpha} - 1 \right) = K , \alpha \neq 1 , \alpha > 0 \tag{48}$$

or $$\sum_{i=1}^{n} p_i^{\alpha} \sum_{j=1}^{k} \lambda_j q_{ij}^{1-\alpha} = K(\alpha-1) + 1 \tag{49}$$

or $$\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} = K(\alpha-1)+1 ; \sum_{j=1}^{k} \lambda_j q_{ij}^{1-\alpha} = q_i^{1-\alpha} \tag{50}$$

or $$\sum_{i=1}^{n} p_i^{\alpha} \tilde{\tilde{q}}_i^{1-\alpha} = A^{\alpha-1}(K(\alpha-1)+1); \tilde{\tilde{q}}_i = \frac{q_i}{A} ; A = \sum_{i=1}^{n} q_i \tag{51}$$

or $$\frac{1}{\alpha-1}\left(\sum_{i=1}^{n} p_i^{\alpha} \tilde{\tilde{q}}_i^{1-\alpha} - 1 \right) = \frac{A^{\alpha-1}(K(\alpha-1)+1)-1}{\alpha-1} \tag{52}$$

which is again a circle of the first type with centre at $\tilde{\tilde{Q}}$ and radius

$[A^{\alpha-1}(K(\alpha-1)+1)-1]/(\alpha-1)$ or $A^{\alpha-1}\left(k + \frac{1-A^{1-\alpha}}{\alpha-1}\right)$.

(b) In this case, the weighted ellipse of the second type is

$$\frac{1}{\alpha-1} \sum_{j=1}^{k} \lambda_j \left(\sum_{i=1}^{n} q_{ij}^{\alpha} p_i^{1-\alpha} - 1 \right) = K, \alpha \neq 1, \alpha > 1 \tag{53}$$

or $$\sum_{i=1}^{n} p_i^{1-\alpha} \sum_{j=1}^{k} \lambda_j q_{ij}^{\alpha} = K(\alpha-1)+1 \tag{54}$$

or $$\sum_{i=1}^{n} p_i^{1-\alpha} \bar{q}_i^{\alpha} = K(\alpha-1)+1 \; ; \; \bar{q}_i^{\alpha} = \sum_{j=1}^{k} \lambda_i q_{ij}^{\alpha} \tag{55}$$

or $$\frac{1}{\alpha-1}\left(\sum_{i=1}^{n} p_i^{1-\alpha} \bar{\bar{q}}_i^{\alpha} - 1 \right) = \frac{B^{\alpha}(K(\alpha-1)+1)-1}{\alpha-1} ,$$

$$\bar{\bar{q}}_i = \frac{\bar{q}_i}{\sum_{i=1}^{n} \bar{q}_i} , \; B = \sum_{i=1}^{n} \bar{q}_i \tag{56}$$

so that the weighted ellipse of the second type is a circle of the second type with centre at $\bar{\bar{Q}}$ and radius $(B^{\alpha}(K(\alpha-1)+1)-1)/(\alpha-1)$.

(c) By proceed ing as in section 5, we can easily shown that for the continuous-variate distributions also, the weighted ellipses of the first and second types are circles of the first and second type, even when Havrda and Charvat's measure of directed divergence is used ,

(d) By maning $\alpha \rightarrow 1$, we can deduce results of sections 2 and 3 as limiting cases of results proved on subsections (a) and (b) above ,

(e) If we use Renyi's [9] measure of directed divergence, the weighted ellipse of the first type is

$$\frac{1}{\alpha-1} \sum_{j=1}^{k} \lambda_j \ln \left(\sum_{i=1}^{n} p_i^{\alpha} q_{ij}^{1-\alpha} \right) = K , \alpha \neq 1, \alpha \geqslant 0 \tag{57}$$

or $$\ln \prod_{j=1}^{k} \left(\sum_{i=1}^{n} p_i^{\alpha} q_{ij}^{1-\alpha} \right)^{\lambda_j} = K(\alpha-1) \tag{58}$$

weich cannot be written normally in the form

$$ln \sum_{i=1}^{n} p_i^{\alpha} \, q_i^{*1-\alpha} = K' \tag{59}$$

Thus, if we use Renyi's measure of directed divergence, the weighted ellipse of the first type is not a circle of ths first type .

(f) Similarly, if we use Kapur's ([3], [4]) or Sharma and Mittal's [11] measures of directed divergence, the weighted ellipses of the first and second type will not in general be circles of the first and second type .

(g) If, for two measures of directed divergence $D_1 (P : Q)$ and $D_2 (P : Q)$, we can write

$$\sum_{j=1}^{k} \lambda_j D_1 (P : Q_j) = A_1 D_1 (P . \bar{Q}) + B_1 , \quad \sum_{j=1}^{k} \lambda_j D_2 (P : Q_j)$$

$$= A_2 D (P : \bar{\bar{Q}}) + B_2 \tag{60}$$

where $\bar{Q}$, $\bar{\bar{Q}}$ are proper probability distributions and $A_1 , A_2 > 0$, then if

$$D (P : Q) = a_1 D_1 (P : Q) + a_2 D_2 (P : Q) , \; a_1 > 0, \; a_2 > 0 \tag{61}$$

we get

$$\sum_{j=1}^{k} \lambda_j D (P : Q_j) = a_1 A_1 D_1 (P : \bar{Q}) + a_2 A_2 D_2 (P : \bar{\bar{Q}}) + a_1 B_1 + a_2 B_2 \tag{62}$$

and the *RHS* of (62) will not in general be of the form $B D(P : \bar{\bar{Q}}) + C$ and as such even if for two measures of directed divergence , $D_1(P : Q)$ and $D_2(P : Q)$, a weighted ellipse becomes a circle, this will not in general happen for a directed divergence measure which is a convex linear combination of $D_1 (P :Q)$ and $D_2 (P : Q)$.

In particular for Sharma and Taneja's [12] measure which is a convex linear combination of Havrda and Charvat's measures, whenever it is valid, a weighted ellipse need not be a circle.

7. CONCLUDING REMARKS

We have proved that when we use Havrda and Charvat's measure of directed divergence or its limiting case viz. Kullback and Leibler's measure of directed divergence, as a measure of distance, then every weighted ellipse whether of first or second type, is a circle of the same type. However this result does not hold for most other measures of directed divergence.

The result will hold for the measure

$$D(P:Q) = \sum_{i=1}^{n} \phi(p_i, q_i) \tag{63}$$

$$\text{if} \quad \sum_{j=1}^{k} \lambda_j \phi(p, q_j) = a\, \phi(p, \bar{q}) + b \tag{64}$$

where $\bar{q}$ is some weighted mean of $q_1, q_2, \ldots, q_k$.

Our other main results concern the minimum and maximum possible values of

$$\sum_{j=1}^{k} \lambda_j D(P:Q_j) \text{ and } \sum_{j=1}^{k} \lambda_j D(Q_j:P) \text{ for variations in } P.$$

(i) For Kullback-Leibler [8] measure $D(P:Q)$

$$\min_P \left[\sum_{j=1}^{k} \lambda_j D(P:Q_j) \right] = -\ln \sum_{i=1}^{n} q_i, \quad \max_P \left[\sum_{j=1}^{k} \lambda_j D(P:Q_j) \right]$$

$$= -\ln \min(q_1, q_2, \ldots, q_n) \tag{65}$$

$$\min_P \left[\sum_{j=1}^{k} \lambda_j D (Q_j : P) \right] = J (\lambda_1 , \lambda_2, \ldots , \lambda_k ; Q_1, Q_2, \ldots , Q_k),$$

$$\max_P \left[\sum_{j=1}^{k} \lambda_j D (Q_j : P) \right] = \infty \qquad (66)$$

$$\min_f \left[\sum_{j=1}^{k} \lambda_j D(f : g_j) \right] = -ln \int g (x) dx ; \max_f \left[\sum_{j=1}^{k} \lambda_j D(f : g_j) \right] = \infty \qquad (67)$$

$$\min_f \left(\sum_{j=1}^{k} \lambda_j D (g_j : f) \right) = J (\lambda_1 , \ldots , \lambda_k ; g_1 , g_2 , \ldots , g_k) ;$$

$$\max_f \left(\sum_{j=1}^{k} \lambda_j D (g_j : f) \right) = \infty \qquad (68)$$

(ii) For Havrda and Charvat's 2-measure of directed divergence

$$\min_P \left[\sum_{j=1}^{k} \lambda_j D (P : Q_j) \right] = \frac{A^{1-\alpha} -1}{\alpha-1} =$$

$$= \frac{\left[\sum_{i=1}^{n} \left(\sum_{j=1}^{k} \lambda_j q_{ij}^{1-\alpha} \right)^{\frac{1}{1-\alpha}} \right]^{1-\alpha} -1}{\alpha-1} \qquad (69)$$

$$\max_P \left[\sum_{j=1}^{k} \lambda_j D (P : Q_j) \right] = \max_i \frac{q_i^{1-\alpha} -1}{\alpha-1} \qquad (70)$$

$$\min_P \left[\sum_{j=1}^{k} \lambda_j D (Q_j : P) = \frac{(\sum_{i=1}^{n} \bar{q}_i)^{-\alpha}}{\alpha-1} \right. . \qquad (71)$$

As $\alpha \to 1$, these give results for Kullback - Leibler's measure [8] .

Similarly expressions can be found for the continuous-variate case .

The RHS of (71) can be regarded as a generalisation of Jensen's difference .

The important points to note is that both for Kullback-Leibler and Havrda and Charvat's measures, both the weighted averages of directed divergences to and from a given set of probability distributions have a minimum value and this result is true for both discrete and continuous-variate distributions ,

The geometry of the probability spacc and some optimization problems therein have been discussed by Kapur ([5] and [6]) .

REFERENCES

[1] J. Burbea and C. R. Rao , On the convexity of some divergence measures based on entropy functions, *IEEE Trans. Inform. Theory* **28** (1982) , 489–495 .

[2] J. Havrda and F. Charvat, Quantification methods of classificatory processes : concept of structural α entropy, *Kybernetika* (*Prague*) 3 (1967), 30–35 .

[3] J. N. Kapur, Generallsed entropy of order α and type β, *Math. Sem.* (*Delhi*) 4 (1967), 78–94 .

[4] J. N. Kapur, Four families of measures of entropy, *Indian J. Pure Appl. Math.* **17** (1986), 429-449 .

[5] J. N. Kapur, *Seven Optimization Problems in Information Theory*, Research Report No. 309, Indian Institute of Technology, Kanpur, 1986 .

[6] J. N. Kapur, *Two Optimization Problems in Probability Space*, Research Report No. 314 Indian Institute of Technology, Kanpur, 1986 .

[7] J. N. Kapur, *Maximum Entropy Models in Science and Engineering*, South Asian Publishers, New Delhi, 1986.

[8] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Statist.* **22** (1951), 79-86.

[9] A. Renyi, On the measures of information and entropy, *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** (1961), 547-561.

[10] C. E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* **27** (1948), 379–423, 623–656.

[11] B. D. Sharma and D. P. Mittal, New non-additive measure of entropy for a discrete probability distribution, *J. Math. Sci.* **10** (1975), 28-40.

[12] B. D. Sharma and I. J. Taneja, Entropy of type α, β and other generalised measures of information theory, *Metrika* **22** (1978), 205-210.