Internat. J. Math. & Math. Sci.
Vol. 10, No.3 (1987) 597-608

A MEASURE OF MUTUAL DIVERGENCE AMONG A NUMBER OF PROBABILITY DISTRIBUTIONS

J.N. KAPUR
Department of Mathematics
Indian Institute of Technology
New Delhi, India 110016

VINOD KUMAR and UMA KUMAR
School of Business
Carleton University
Ottawa, Canada K1S 5B6

(Received September 11, 1985 and in revised form September 22, 1986)

ABSTRACT. The principle of optimality of dynamic programming is used to prove three major inequalities due to Shannon, Renyi and Holder. The inequalities are then used to obtain some useful results in information theory. In particular measures are obtained to measure the mutual divergence among two or more probability distributions.

KEY WORDS AND PHRASES. Divergence measures, dynamic programming, Shannon, Renyi and Holder inequalities.

1980 AMS SUBJECT CLASSIFICATION CODE. 94A17, 94A15, 90C50, 90C39

1. INTRODUCTION

An important problem in business, economic and social sciences is concerned with measuring differences in specific characteristics of business, economic or social groups. For example, we may like to compare the balance sheets of m business firms, or economic development of m nations or the difference in opinions of m directors of a company or the population compositions of m cities and so on.

Let q_{ij} represent the share of the jth item from the ith group. Thus q_{ij} may represent proportionate contributions of the jth item in the balance sheet of the ith company or it may represent the population of persons in the jth income group of the ith nation or the proportionate allotment to the jth item of the budget proposed by the ith director or the proportionate population of the jth social group in the ith city and so on. Let

$$Q_i = (q_{i1}, q_{i2}, \cdots, q_{in});\ i = 1, 2, \cdots, m$$

where

$$\sum_{i=1}^{n} q_{ij} = 1, \quad q_{ij} \geq 0$$

Given $Q_1, Q_2, \cdots, Q_m$, our problem is to find how different $Q_1, Q_2, \cdots, Q_m$ are. We can regard $Q_1, Q_2, \cdots, Q_m$ as m probability distributions and thus our problem is to find a measure of mutual divergence between $Q_1, Q_2, \cdots, Q_m$.

This measure should have the following properties:

(i) It should depend on $Q_1, Q_2, \cdots, Q_m$

(ii) It should be a continuous function of all q_{ij}'s

(iii) It should not change when $q_{i1}, q_{i2}, \cdots, q_{in}$ are permuted among themselves, provided the same permutation is used for each of $Q_1, Q_2, \cdots, Q_m$.

(iv) It should be always ≥ 0

(v) It should vanish when $Q_1 = Q_2 = \cdots = Q_m$

(vi) It should vanish only when $Q_1 = Q_2 = \cdots = Q_m$

(vii) It should possibly be a convex function of $Q_1, Q_2, \cdots, Q_m$

For two distributions, ($m = 2$) we have a number of measures of directed divergence due to Kullback-Leibler, Havrda-Charvat, Renyi, Kapur and others [5]. However, in real life we are concerned with $m (\geq 2)$ distributions. We can of course find directed divergence between every pair of m distributions and then find some sort of an average of these directed divergences.

However, we shall prefer to have a unified measure depending directly on the m distributions. We develop such a measure in the present paper.

Since one important condition for the measure is non-negativity and this condition is expressed as an inquality, an important role is played in the development of our new measure by the special inequalities due to Shannon, Renyi, Holder and their generalizations. We give alternate proofs of all these inqualities by using dynamic programming and then use these inequalities to develop our new measure.

2. DYNAMIC PROGRAMMING AND INEQUALITIES

Kapur [4] used dynamic programming technique of Bellman [2] to show that the maximum value of $\sum_{i=1}^{n} p_i \ln p_i$ subject to $p_1 \geq 0, p_2 \geq 0, 0, \cdots, p_n \geq 0$, $\sum_{i=1}^{n} p_1 = c$ is $-c \ln(c/n)$ and it arises when $p_1 = p_2 = \cdots = p_n = c/n$. In particular, when $c = 1$, we find that the maximum value of the Shannon's [7] entropy measure $-\sum_{i=1}^{n} p_i \ln p_i$ subject to $p_1 \geq 0$, $p_2 \geq 0, \cdots, p_n \geq 0$, $\sum_{i=1}^{n} p_i = 1$ is $\ln n$ and arise when all the probabilities are equal. The result is equivalent to showing that

$$-\sum_{i=1}^{n} p_i \ln p_i \leq \ln n \tag{2.1}$$

for all probability distributions.

The same technique can be used to establish inequalities of the form

$$\sum_{i=1}^{n} f(p_i, q_i) \leq A \text{ or } \sum_{i=1}^{n} f(p_i, q_i) \geq B \tag{2.2}$$

In the first case, we show that the maximum value of $\sum_{i=1}^{n} f(p_i, q_i)$ is A and in the second case, we show that the minimum value of $\sum_{i=1}^{n} f(p_i, q_i)$ is B.

Suppose we have to maximize $\sum_{i=1}^{n} f(p_i, q_i)$ subject to

$$p_1 \geq 0, p_2 \geq 0, \cdots, p_n \geq 0;\ q_1 \geq 0,\ q_2 \geq 0, \cdots, q_n \geq 0,\ \sum_{i=1}^{n} p_i = a, \sum_{i=1}^{n} q_i = b \tag{2.3}$$

Let the maximum value be $g_n(a, b)$.

We may keep $p_1, p_2, \cdots, p_n$ fixed‘and vary only $q_1, q_2, \cdots, q_n$ subject to $\sum_{i=1}^{n} q_i = b$, then the principle of optimality gives

$$g_n(a, b) = \max_{0 \leq q_1 \leq b} [f(p_1, q_1) + g_{n-1}(a - p_1, b - q_1)] \tag{2.4}$$

Since p_1 is fixed, we have to maximize a function of one variable q_1 only. Equation (2.4) give a recurrence relation which along with the knowledge of $f_1(a, b)$ can enable us to obtain $g_n(a, b)$ for all values of n, by proceeding step by step.

The same procedure can be used to find the minimum value.

3. SHANNON'S INEQUALITY

Here

$$f(p_i, q_i) = p_i \ln \frac{p_i}{q_i} \tag{3.1}$$

This is a convex function of q_i and we seek to determine the minimum value of $\sum_{i=1}^{n} f(p_i, q_i)$. We get

$$g_1(a, b) = a \ln \frac{a}{b} \tag{3.2}$$

Using (2.4)

$$\begin{aligned} g_2(a, b) &= \max_{0 \leq q_1 \leq b} \left[p_1 \ln \frac{p_1}{q_1} + (a - p_1) \ln \frac{a - p_1}{b - q_1}\right] \\ &= a \ln \frac{a}{b} \end{aligned} \tag{3.3}$$

Assuming

$$g_n(a, b) = a \ln \frac{a}{b}, \tag{3.4}$$

(2.4) gives

$$\begin{aligned} g_{n+1}(a, b) &= \min_{0 \leq q_1 \leq b} \left[p_1 \ln \frac{p_1}{q_1} + (a - p_1) \ln \frac{a - p_1}{b - q_1}\right] \\ &= a \ln \frac{a}{b} \end{aligned} \tag{3.5}$$

Hence by mathematical induction we get

$$\min \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} = a \ln \frac{a}{b} \tag{3.6}$$

or

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \geq a \ln \frac{a}{b} \tag{3.7}$$

The minimum value is obtained when

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \cdots = \frac{p_n}{q_n} = \frac{a}{b} \tag{3.8}$$

If $a = b = 1$, then we get Shannon's inequality

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \geq 0 \text{ and } \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} = 0 \text{ iff } p_i = q_i \text{ for each } i \tag{3.9}$$

If $b \leq a$,

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \geq a \ln \frac{a}{b} \geq 0 \tag{3.10}$$

Alternative proofs are given in Aczel and Daroczy [1].

4. RENYI'S INEQUALITY

Here

$$f(p_i, q_i) = \frac{1}{\alpha - 1}(p_i^{\alpha} q_i^{1-\alpha} - p_i) \tag{4.1}$$

$$g_1(a, b) = \frac{1}{\alpha - 1}(a^{\alpha} b^{1-\alpha} - a) \tag{4.2}$$

$f(p_i, q_i)$ is a convex function of q_i and we try to find the minimum value of $\sum_{i=1}^{n}(p_i, q_i)$

$$\begin{aligned} g_2(a, b) &= \min_{0 \leq q_1 \leq b} \left[\frac{1}{\alpha - 1}(p_1^{\alpha} q_1^{1-\alpha} - p_1) + \frac{1}{\alpha - 1}(a - p_1)^{\alpha}(b - q_1)^{1-\alpha} - (a - p_1)\right] \\ &= \frac{1}{\alpha - 1}(a^{\alpha} b^{1-\alpha} - a). \end{aligned} \tag{4.3}$$

By using mathematical induction, we can show that

$$\frac{1}{\alpha - 1}\left[\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} - \sum_{i=1}^{n} p_i\right] \geq \frac{1}{\alpha - 1}\left[a^{\alpha} b^{1-\alpha} - a\right] \tag{4.4}$$

If $a = b = 1$, we get Renyi's inequality

$$\frac{1}{\alpha - 1}\left[\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} - 1\right] \geq 0 \tag{4.5}$$

The inequality (4.5) will hold whenever $a = b$, even if the common value is not unity.

Alternative proofs of this inequality are available in Aczel and Daroczy [1] and Kapur [6].

5. HOLDER'S INEQUALITY

Let $f_n(M)$ denote the minimum value of

$$U = \left(\sum_{j=1}^{n} x_j^p\right)^{q/p} \left(\sum_{j=1}^{n} y_j^q\right), \tag{5.1}$$

where x_j's and y_j's are positive real numbers, x_j's are fixed and y_j's vary subject to $\sum_{j=1}^{n} x_j y_j = M$, then

$$f_1(M) = M^q \tag{5.2}$$

$$f_2(M) = (x_1^p + x_2^p)^{q/p} \min_{0 \le z_1 \le M} \left[\left(\frac{z_1}{x_1}\right)^q + \left(\frac{M - z_1}{x_2}\right)^q\right] \tag{5.3}$$

Now

$$\begin{aligned} g(z_1) = \left(\frac{z_1}{x_1}\right)^q + \left(\frac{M - z_1}{x_2}\right)^q &\Rightarrow g'(z_1) = q\left(\frac{z_1^{q-1}}{x_1^q} - \frac{(M - z_1)^{q-1}}{x_2^q}\right) \\ &\Rightarrow g''(z_1) = q(q-1)\left(\frac{z_1^{q-2}}{x_1^q} + \frac{(M - z_1)^{q-2}}{x_2^q}\right) \end{aligned} \tag{5.4}$$

If $q(q-1) > 0$, the minimum value of $g(z_1)$ occurs when

$$\frac{z_1}{x_1^{q/q(q-1)}} = \frac{M - z_1}{x_2^{q/(q-1)}} = \frac{M}{x_1^{q/(q-1)} + x_2^{q/(q-1)}} \tag{5.5}$$

and

$$[g(z_1)]_{min} = M^q / \left[x_1^{q/(q-1)} + x_2^{q/(q-1)}\right]^{q-1} \tag{5.6}$$

$$= M^q / [x_1^p + x_2^p]^{q-1} \tag{5.7}$$

If

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \text{or} \quad q - 1 = q/p, \tag{5.8}$$

then

$$f_2(M) = (x_1^p + x_2^p)^{q/p} \frac{M^q}{(x_1^p + x_2^p)^{q/p}} = M^q \tag{5.9}$$

If we assume that $f_n(M) = M^q$, then the principle of optimality gives

$$\begin{aligned} f_{n+1}(M) &= \left(\sum_{i=1}^{n} x_i^p\right)^{q/p} \min_{0 \le z_1 \le M} \left[\left(\frac{z_1}{x_1}\right)^q + \frac{(M - z_1)^q}{(x_n^p + \cdots + x_{n+1}^p)^{q/p}}\right] \\ &= M^q \end{aligned} \tag{5.10}$$

thus if $q(q-1) > 0$,

$$\left(\sum_{j=1}^{n} x_j^p\right)^{q/p} \left(\sum_{j=1}^{n} y_j^q\right) \ge M^q \tag{5.11}$$

If $q > 1$, this gives

$$\sum_{j=1}^{n} x_j y_j \le \left(\sum_{j=1}^{n} x_j^p\right)^{1/p} \left(\sum_{j=1}^{n} y_j^q\right)^{1/q} \tag{5.12}$$

Using (5.10), we find that the inequality (5.12) holds if $q > 1, p > 1, p^{-1} + q^{-1} = 1$.

If $q < 0, 0 < p < 1$ or $o < q < 1$ and $p < 0$ and $p^{-1} + q^{-1} = 1$, then the inequality in (5.12) is reversed and we get

$$\sum_{j=1}^{n} x_j y_j \leq \left(\sum_{j=1}^{n} x_j^p\right)^{1/p} \left(\sum_{j=1}^{n} y_j^q\right)^{1/q} . \tag{5.13}$$

6. A GENERALIZATION OF HOLDER'S INEQUALITY

Now we consider the minimization of

$$V_n = \left(\sum_{j=1}^{n} x_j^p\right)^{r/p} \left(\sum_{j=1}^{n} y_j^q\right)^{r/q} \left(\sum_{j=1}^{n} z_j^r\right), \tag{6.1}$$

where x_j's, y_j's, z_j's are all positive real numbers, x_j's and y_j's are kept fixed and z_j's vary subject to

$$\sum_{j=1}^{n} x_j y_j z_j = M \tag{6.2}$$

If $k_n(M)$ is the minimum value of U, then

$$k_1(M) = M^r \tag{6.3}$$

$$\begin{aligned} k_2(M) &= \left(\sum_{j=1}^{2} x_j^p\right)^{r/p} \left(\sum_{j=1}^{2} y_j^q\right)^{r/q} \min_{0 \leq u \leq M} \left[\left(\frac{u}{x_1 y_1}\right)^r + \left(\frac{M-u}{x_2 y_2}\right)^r\right] \\ &= \left(\sum_{j=1}^{2} x_j^p\right)^{r/p} \left(\sum_{j=1}^{2} y_j^q\right)^{r/p} \left[\frac{M^r}{(x_1 y_1)^{r/(r-1)} + (x_2 y_2)^{r/(r-1)}}\right] \end{aligned} \tag{6.4}$$

provided $r > 1$, so that in this case

$$V_2 \geq M^r \left(\sum_{j=1}^{2} x_j^p\right)^{r/p} \left(\sum_{j=1}^{2} y_j^q\right)^{r/q} \left[(x_1 y_1)^{r/(r-1)} + (x_2 y_2)^{r/(r-1)}\right]^{-(r-1)} \tag{6.5}$$

Similarly when $r > 1$

$$V_n \geq M^r \left(\sum_{j=1}^{n} x_j^p\right)^{r/p} \left(\sum_{j=1}^{n} y_j^q\right)^{r/q} \left[\sum_{j=1}^{n} (x_j y_j)^{r/(r-1)}\right]^{-(r-1)} \tag{6.6}$$

Now we consider the case when $p > 1, q > 1, r > 1, p^{-1} + q^{-1} + r^{-1} = 1$, so that

$$\frac{1}{p} + \frac{1}{q} = \frac{r-1}{r} \text{or } \frac{r}{r-1}\frac{1}{p} + \frac{r}{r-1}\frac{1}{q} = 1$$

or

$$\frac{1}{p'} + \frac{1}{q'} = 1;\ p' = p\frac{r-1}{r},\ q' = q\frac{r-1}{r} \tag{6.7}$$

Again by Holder's inequality proved in the last section when $p' > 1, q' > 1$

$$\left(\sum_{j=1}^{n} u_j^{p'}\right)^{1/p'} \left(\sum_{j=1}^{n} v_j^{q'}\right)^{1/q'} \geq \sum_{j=1}^{n} u_j v_j$$

or

$$\left(\sum_{j=1}^{n} u_j^{\frac{p(r-1)}{r}}\right)^{\frac{r}{p(r-1)}} \left(\sum_{j=1}^{n} v_j^{\frac{p(r-1)}{r}}\right)^{\frac{r}{q(r-1)}} \geq \sum_{j=1}^{n} u_j v_j$$

or

$$\left(\sum_{j=1}^{n} x_j^p\right)^{r/p} \left(\sum_{j=1}^{n} y_j^q\right)^{r/q} \geq \left(\sum_{j=1}^{n} x_j^{\frac{r}{r-1}} y_j^{\frac{r}{r-1}}\right)^{r-1} \tag{6.8}$$

From (6.6) and (6.8), $V_n \geq M^r$ or

$$\left(\sum_{j=1}^{n} x_j^p\right)^{1/p} \left(\sum_{j=1}^{n} y_j^q\right)^{1/q} \left(\sum_{j=1}^{n} z_j^r\right)^{1/r} \geq \sum_{j=1}^{n} x_j y_j z_j \tag{6.9}$$

whenever (6.7) is satisfied.

The method of proof is quite general and (6.9) is easily generalized to give

$$\left(\sum_{j=1}^{n} x_{j1}^{q1}\right)^{1/q_1} \left(\sum_{j=1}^{n} x_{j2}^{q2}\right)^{1/q_2} \cdots \left(\sum_{j=1}^{n} x_{jm}^{q_m}\right)^{1/q_m} \geq \sum_{j=1}^{n} x_{j1} x_{j2} \cdots x_{jm} \tag{6.10}$$

whenever

$$q_1 > 1, q_2 > 1, \cdots, q_m > 1;\ q_1^{-1} + q_2^{-1} + \cdots + q_m^{-1} = 1 \tag{6.11}$$

Inequality (6.10) can also be written as

$$\sum_{j=1}^{n} y_{j1}^{\alpha_1} y_{j2}^{\alpha_2} \cdots y_{jm}^{\alpha_m} \leq \left(\sum_{j=1}^{n} y_{j1}\right)^{\alpha_1} \left(\sum_{j=1}^{n} y_{j2}\right)^{\alpha_2} \cdots \left(\sum_{j=1}^{n} y_{jm}\right)^{\alpha_m} \tag{6.12}$$

whenever

$$0 < \alpha_1 < 1, 0 < \alpha_2 < 2, \cdots, 0 < \alpha_m < 1;\ \alpha_1 + \alpha_2 + \cdots + \alpha_m = 1. \tag{6.13}$$

If y_{ji}'s represent probabilities, we get

$$\sum_{j=1}^{n} p_{j1}^{\alpha_1} p_{j2}^{\alpha_2} \cdots p_{jm}^{\alpha_m} \leq 1 \tag{6.14}$$

In particular for $m = 2$, (6.14) gives

$$\sum_{j=1}^{n} p_j^{\alpha} q_j^{1-\alpha} \leq 1 \quad \text{when } 0 < \alpha < 1$$

or

$$(\alpha - 1)^{-1} \left(\sum_{j=1}^{n} p_j^{\alpha} q_j^{1-\alpha} - 1\right) \geq 0 \tag{6.15}$$

Alternative proofs of Holder's inequality are given in Hardy, Littlewood and Polya [3] and Marshall and Olkin [7].

7. HOLDER'S GENERALIZED INEQUALITY FOR OTHER VALUES OF PARAMETERS

Holder's generalized inequality (6.10) holds whenever $q_1, q_2 \cdots, q_n$ satisfy (6.11). However, if (6.11) are not satisfied, the inequality sign in (6.10) may be reversed or no inequality may hold.

Thus for the special case $m = 3$ if $0 < r < 1$, the inequality sign in (6.5) is reversed. If one of p or q is negative, then in Holder's inequality also the sign is reversed so that if $0 < r < 1$, one of p and q is negative and $p^{-1} + q^{-1} + r^{-1} = 1$, then

$$\left(\sum_{j=1}^{n} x_j^p\right)^{1/p} \left(\sum_{j=1}^{n} y_j^q\right)^{1/q} \left(\sum_{j=1}^{n} z_j^r\right)^{1/r} \leq \sum_{j=1}^{n} x_j y_j z_j \tag{7.1}$$

while if $r > 1, p > 1, q > 1$ and $p^{-1} + q^{-1} + r^{-1} = 1$, then inequality (6.9) holds. If $p^{-1} + q^{-1} + r^{-1} = 1$, but neither of the two sets of conditions is satisfied, then neither (6.9) nor (7.1) may hold. Thus

if $p = 2, q = 4, r = 4$ (6.9) holds

if $p = 1/2, q = -1/3, r = 1/2$, (7.1) holds

if $p = -2, q = -2, r = 1/2$, neither (6.9) nor (7.1) holds.

Similarly given $q_1, q_2, \cdots, q_n$ we can find whether (6.10) holds or whether (6.10) with reversed sign of inequality holds or whether neither of these two inequalities holds.

8. APPLICATIONS TO INFORMATION THEORY

From Renyi's or Holder's inequality, it follows that

$$D_\alpha(P : Q) = \frac{1}{\alpha - 1}\left(\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} - 1\right) \geq 0, \alpha \neq 1, \alpha > 0 \tag{8.1}$$

where $P = (p_1, p_2, \cdots, p_n)$, $Q = (q_1, q_2, \cdots, q_n)$ are two probability distributions. Also $D_\alpha(P : Q) = 0$ iff $p_i = q_i$ for each i. In the limiting case when $\alpha \to 1$, this gives

$$D_1(P : Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \geq 0 \tag{8.2}$$

$D_\alpha(P : Q)$ or $D_1(P : Q)$ can be used as a measure of directed divergence of P from Q.

for these distributions, we similarly have from generalized Holder's inequality, that

$$D_{\alpha,\beta}(P : Q : R) = \frac{1}{\alpha^2 + \beta^2 - 1}\left[\sum_{i=1}^{n} p_i^\alpha q_i^\beta r_i^{1-\alpha-\beta} - 1\right] \geq 0 \tag{8.3}$$

whenever $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, $\alpha + \beta < 1$, and it vanishes iff $p_i = q_i = r_i$ for each i.

If $\alpha = 0$, it measures the directed divergence of Q from P

$\beta = 0$, it measures the directed divergence of P from R and if

$\alpha + \beta = 1$, it measures the directed divergence of P from Q.

Collectively $D_{\alpha,\beta}(P : Q : R)$ can be used as a measure of mutal divergence among three distributions.

Generalizing we get the measure

$$\begin{aligned} &D_{\alpha_1,\alpha_2,\cdots,\alpha_m}(P_1 : P_2 : \cdots : P_m) \\ &= \frac{1}{\alpha_1^2+\alpha_2^2+\cdots\alpha_m^2-1}\left[\sum_{i=1}^{n} p_{i1}^{\alpha_1} p_{i2}^{\alpha_2}\cdots p_{im}^{\alpha_m}-1\right] \geq 0 \end{aligned} \tag{8.4}$$

where $0<\alpha_1<1, 0<\alpha_2<1,\cdots,0<\alpha_m<1$, $\alpha_1+\alpha_2+\cdots\alpha_m=1$. By putting some of $\alpha_1,\alpha_2,\cdots,\alpha_m$ equal to zero, it can also be used to measure all divergences among subsets of $P_1,P_2,\cdots,P_m$.

for three distributions, Theil [9] had proposed the measure

$$D(P:Q:R)=\sum_{i=1}^{n} p_i \ln\frac{r_i}{q_i} \tag{8.5}$$

as a measure of information improvement. However, this can be both negative and positive. On the other hand our measure (8.3) is always ≥ 0 and can be a more effective measure of improvement.

The measure (8.3) and (8.4) are also recursive since

$$D_{\alpha,\beta,n}(P:Q:R)=\frac{1}{\alpha^2+\beta^2-1}\left[\sum_{i=1}^{n} p_i^{\alpha} q_i^{\beta} r_1^{1-\alpha-\beta}-1\right]$$

or

$$\begin{aligned} (\alpha^2+\beta^2-1)D_{\alpha,\beta,n}(P:Q:R) &= p_1^{\alpha}q_1^{\beta}r_1^{1-\alpha-\beta}+p_2^{\beta}q_2^{\beta}r_2^{1-\alpha-\beta} \\ &\quad-(p_1+p_2)^{\alpha}(q_1+q_2)^{\beta}(r_1+r_2)^{1-\alpha-\beta} \\ &\quad+(\alpha^2+\beta^2-1)D_{\alpha,\beta,n-1}(p_1,p_2,p_3,\cdots,p_n; \\ &\quad q_1+q_2,q_3,\cdots,q_n;\ r_1+r_2,r_3,\cdots,r_n) \end{aligned} \tag{8.6}$$

or

$$\begin{aligned} &D_{\alpha,\beta,n}(p_1,\cdots,p_n;q_1,\cdots,q_n;r_1,\cdots,r_n) \\ &= D_{\alpha,\beta,n-1}(p_1+p_2,p_3,\cdots,p_n;q_1+q_2,q_3,\cdots,q_n;r_1+r_2,r_3,\cdots,r_n) \\ &\quad+(p_1+p_2)^{\alpha}(q_1+q_2)^{\beta}(r_1+r_2)^{1-\alpha-\beta}. \\ &H_2\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2},\frac{q_1}{q_1+q_2},\frac{q_2}{q_1+q_2},\frac{r_1}{r_1+r_2},\frac{r_2}{r_1+r_2}\right) \end{aligned} \tag{8.7}$$

9. MEASURE OF SYMMETRIC MUTUAL DIVERGENCE

The measure (8.3) is not symmetric in the sense that

$$D_{\alpha,\beta}(P:Q:R), D_{\alpha,\beta}(Q:P:R), D_{\alpha,\beta}(P:R:Q)$$

etc. are not equal unless $\alpha=\beta=1/3$. The measure

$$D_{\frac{1}{3},\frac{1}{3}}(P:Q:R)=\frac{9}{7}\left[1-\sum_{i=1}^{n} p_i^{1/3}q_i^{1/3}r_i^{1/3}\right] \tag{9.1}$$

is symmetric. Another symmetric measure of mutal divergence is given by

$$\begin{aligned} &D_{\alpha,\beta}(P:Q:R)+D_{\alpha,\beta}(P:R:Q)+D_{\alpha,\beta}(Q:P:R) \\ &\qquad+D_{\alpha,\beta}(Q:R:P)+D_{\alpha,\beta}(R:Q:P) \end{aligned} \tag{9.2}$$

The corresponding symmetric measure for m distributions will be the sum of $m!$ mutual divergences.

This is a two-parameter family of measures of mutal symmetric divergences.

The measure (9.2) can be compared with the measure

$$D_\alpha(P:Q)+D_\alpha(Q:P)+D_\alpha(P:R)+D_\alpha(R:P)+D_\alpha(Q:R)+D_\alpha(R:Q) \tag{9.3}$$

which is also a measure of symmetric mutal divergence.

10. CONCLUDING REMARKS

By using dynamic programming, it can be easily shown that the maximum value of $x_1, x_2 \cdots x_n$ subject to $x_1 + x_2 + \cdots + x_n = c$, $x_1 \geq 0, \cdots, x_n \geq 0$ is $(c/n)^n$ and the minimum value of $x_1 + x_2 + \cdots + x_n$ subject to $x_1 x_2 \cdots x_n = 1$, $x_1 > 0, x_2 > 0, \cdots, x_n > 0$ is $n^n\sqrt{d}$. From either of these results, the Arithmetic-Geometric-Mean Inequality viz.

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt{x_1 x_2 \cdots x_n} \tag{10.1}$$

can be deduced. From this inequality, one can obtain Shannon's and Holder's inequalities. One can also prove Renyi's entropy first by using dynamic programming and then deduce Shannon's inequality from it. One can also prove first Holder's inequality, by using dynamic programming and then deduce Renyi's Shannon's and AGM inequalities from it. In a sense all these inequalities are equivalent.

Holder's, Renyi's and Shannon's inequality enable us to get a number of measure of directd divergence between two given distributions and with the help of these measures of directed divergence. We can obtain a number of measures of entropy and inaccuracy.

The generalized Holder's inequality enable us to get a number of measures of mutal divergence between $m(> 2)$ probability distributions.

ACKNOWLEDGEMENT

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. ACZEL, G. and DAROCZY, Z. On Measures of Information and Their Characterization, Academic Press, New York, 1975.
2. BELLMAN, R. and DREYFUS, S.E., Applied Dynamic Programming, 1962.
3. HARDY, G.H., LITTLEWOOD and POLYA, G. Inequalities, Cambridge University Press, 1952.

4. KAPUR, J.N. On Some Applications of Dynamic Programming to Information Theory, Proc. Ind. Acad. Sci. , 68A (1968) 1-11.
5. KAPUR, J.N. A Comparative Assessment of Measure of Directed Divergence, Adv. Manag. Studies , 5 (1984) 1-16.
6. KAPUR, J.N. Convex Semi-Metric Spaces and Generalized Shannon's Inequalities. Ind. Journ. Pure Appl. Math. , (to appear), 1986.
7. MARSHALL, A.W., and OLKIN, I. Inequalities Theory of Majorisation and its Applications, Academic Press, New York, 1979.
8. SHANNON, C.E. A Mathematical Theory of Communication, Bell System Technical J. , 27 (1948), 379-423, 623-656.
9. THEIL, H. Economics and Information Theory, North Holland, Amsterdam, 1967.