ON INFORMATION IMPROVEMENT

by RAMINDER SINGH SONI*, *Department of Mathematics, University of Delhi, Delhi* 110007 (*India*)

(*Received* 4 *September* 1974)

The paper presents a characterization of information improvement, viz.,

$$I(P\|Q\|R) = \sum_{i=1}^{m} p_i \log (r_i/q_i),$$

obtained by revising a distribution $Q = (q_1, q_2, \cdots, q_m)$ as $R = (r_1, r_2, \cdots, r_m)$ on the basis of realization $P = (p_1, p_2, \cdots, p_m)$ about a discrete random variate. In the process, the most general continuous solution of a functional equation in function of three variables has been obtained. A situation under which this measure is non-negative has also been discussed. The effect of partitioning the set of events is also examined.

1. INTRODUCTION

Chaundy and McLeod (1960) considered a functional equation

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i, y_j) = \sum_{i=1}^{m} f(x_i) + \sum_{j=1}^{n} f(y_j), \qquad \ldots(1.1)$$

where

$$x_i, y_j \geqslant 0, \sum_{i=1}^{m} x_i = 1 = \sum_{j=1}^{n} y_j,$$

and in terms of its continuous solution given by

$$f(x) = Cx \log_a x,\ a > 1 \qquad \ldots(1.2)$$

it gives Shannon's entropy (1948) of a distribution $P = (p_1, p_2, \ldots, p_m)$, viz.,

$$H(P) = \sum_{i=1}^{m} f(p_i) = - \sum_{i=1}^{m} p_i \log_2 p_i, \qquad \ldots(1.3)$$

when $f(\frac{1}{2}) = 1$.

Later, Kannappan (1972) studied the solutions of the functional equation

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i y_j, u_i v_j) = \sum_{i=1}^{m} f(x_i, u_i) + \sum_{j=1}^{n} f(y_j, v_j), \qquad \ldots(1.4)$$

where $x_i, y_j, u_i, v_j \geqslant 0$, $\sum_{i=1}^{m} x_i = 1 = \sum_{j=1}^{n} y_j$, $\sum_{i=1}^{m} u_i \leqslant 1$ and $\sum_{j=1}^{n} v_j \leqslant 1$, under suitable conditions to give characterizations of Kerridge's inaccuracy (1961) and Kullback's relative information (1959).

---

* Present address: S. G. T. B. Khalsa (E) College, New Delhi 110005.

If $P = (p_1, p_2, \ldots, p_m)$, $\sum_{i=1}^{m} p_i = 1$ and $Q = (q_1, q_2, \ldots, q_m)$, $\sum_{i=1}^{m} q_i = 1$ are two probability distributions of a discrete random variate, then measures of inaccuracy and relative information (directea divergence) are given respectively by

$$H_I(P/Q) = -\sum_{i=1}^{m} p_i \log q_i \qquad \ldots(1.5)$$

and

$$I_D\,(P/Q) = \sum_{i=1}^{m} p_i \log (p_i/q_i). \qquad \ldots(1.6)$$

There is yet another very useful measure in information theory. It measures improvement in information obtained by revising the prediction probability distribution $Q = (q_1, q_2, \ldots, q_m)$ as $R = (r_1, r_2, \ldots, r_m)$ on the basis of the realization $P = (p_1, P_2, \ldots, p_m)$. Theil's expression (1967) for information improvement is given by

$$I(P \parallel Q \parallel R) = \sum_{i=1}^{m} p_i \log (r_i/q_i). \qquad \ldots(1.7)$$

This has been studied differently by Sharma and Ram Autar (1973).

In this paper, we shall characterize this measure. In the process, we obtain the most general continuous solution of the functional equation

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i y_j, u_i v_j, p_i q_j) = \sum_{i=1}^{m} f(x_i, u_i, p_i) + \sum_{j=1}^{n} f(y_j, v_j, q_j), \qquad \ldots(1.8)$$

where

$$x_i, y_j, u_i, v_j, p_i, q_j > 0,\ \sum_{i=1}^{m} x_i = 1 = \sum_{j=1}^{n} y_j,\ \sum_{i=1}^{m} u_i \leqslant 1,$$

$$\sum_{j=1}^{n} v_j \leqslant 1,\quad \sum_{i=1}^{m} p_i \leqslant 1 \quad \text{and} \quad \sum_{j=1}^{n} q_j \leqslant 1.$$

In another section, handling a problem of maximization through dynamic programming, we have studied the condition under which $I(P \parallel Q \parallel R)$ is non-negative. A result on partitioning the events has also been obtained.

2. Most General Continuous Solution of the Functional Equation

We first obtain the most general continuous solution of the functional equation (1.8).

Theorem 1—The most general continuous solution defined on $J = [0, 1) x [0, 1] x (0, 1]$ of the functional equation (1.8) under the condition

$$f(1, 1, 1) = 0 \qquad \ldots(2.1)$$

is given by

$$f(x, y, z) = Ax \log (z/x) + Bx \log (y/x) + Cx \log x, \qquad ...(2.2)$$

where A, B and C are arbitrary constants.

PROOF: At first, let us set

$$x_1 = 1 = y_1 = v_1 = q_1, \quad x_2 = \ldots = x_m = 0 = y_2 = \ldots = y_m$$
$$= v_2 = \ldots = v_m = q_2 = \ldots = q_m, \quad u_1 = \tfrac{1}{2} = u_2 = p_1 = p_2,$$
$$u_3 = \ldots = u_m = 0 = p_3 = \ldots = p_m.$$

The functional equation (1.8), in view of (2.1), gives

$$f(0, 0, 0) = 0. \qquad ...(2.3)$$

Next, put $p_i = u_i = x_i$ and $q_j = v_j = y_j$ and take $f(x_i, x_i, x_i) = f(x_i)$. This reduces (1.8) to (1.1) and thus

$$f(x, x, x) = f(x) = Cx \log x, \text{ for } x \in [0, 1] \qquad ...(2.4)$$

where C is an arbitrary constant. Eqn. (2.4) is the most general continuous form when all the three variables become equal.

Now, putting

$$x_1 = x, x_2 = 1 - x, \quad u_1 = u, \quad u_2 = w, \quad p_1 = p, \quad p_2 = t, \quad u + w \leqslant 1,$$
$$p + t \leqslant 1, \quad x_3 = \ldots = x_m = 0 = u_3 = \ldots = u_m = p_3 = \ldots = p_m,$$

and

$$y_1 = 1, \quad v_1 = v, \quad q_1 = q, \quad v \leqslant 1, \quad q \leqslant 1, \quad y_2 = \ldots = y_m = 0$$
$$= v_2 = \ldots = v_m = q_2 = \ldots = q_m,$$

in (1.8) and using (2.3), we get

$$f(x, uv, pq) + f(1 - x, wv, tq)$$
$$= f(x, u, p) + f(1 - x, w, t) + f(1, v, q), \qquad ...(2.5)$$

for $x, u, p, w, t \in [0, 1]$, $v, q \in (0, 1]$; (x, u, p), $(1-x, w, t) \in J$. First, letting $p = u = x$, $t = w = 1 - x$ in (2.5) and using (2.4), we see that

$$f(x, xv, xq) + f[(1 - x), (1 - x) v, (1 - x) q]$$
$$= Cx \log x + C(1 - x) \log (1 - x) + f(1, v, q), \qquad ...(2.6)$$
$$\text{for } x \in [0, 1], v, q \in (0, 1].$$

Then, by taking $t = w = 1 - x$ in (2.5) and using (2.4), we get

$$f(x, uv, pq) + f[(1 - x), (1 - x) v, (1 - x) q]$$
$$= f(x, u, p) + C(1 - x) \log (1 - x) + f(1, v, q), \qquad ...(2.7)$$

for $u + 1 - x \leqslant 1$ and $p + 1 - x \leqslant 1$, i.e., $u, p \leqslant x$, $x \in [0, 1]$ and $v, q \in (0, 1]$.

Now, (2.6) and (2.7) yield

$$f(x, uv, pq) = f(x, u, p) + f(x, xv, xq) - Cx \log x, \qquad \ldots(2.8)$$

for $u, p \leqslant x$, $x \in [0, 1]$, $v, q \in (0, 1]$.

Let us now define

$$\phi_x(u, p) = f(x, ux, px) - Cx \log x,$$
$$\text{for } x \in [0, 1], u, p \in (0, 1] \qquad \ldots(2.9)$$

Replacing u by ux and p by px in (2.8) and using (2.9), we get

$$\phi_x(uv, pq) = \phi_x(u, p) + \phi_x(v, q), \text{ for } u, v, p, q \in (0, 1] \qquad \ldots(2.10)$$

Taking $u = v = 1$ in (2.10), we get

$$\phi_x(1, pq) = \phi_x(1, p) + \phi_x(1, q), \text{ for } p, q \in (0, 1]. \qquad \ldots(2.11)$$

Equation (2.11) is Cauchy's functional equation; its most general continuous solution (Aczel 1966) is given by

$$\phi_x(1, p) = A(x) \log p, \text{ where } A(x) \text{ depends upon } x \text{ for } p \in (0, 1]. \qquad \ldots(2.12)$$

Similarly, by taking $p = q = 1$ in (2.10), we have

$$\phi_x(u, 1) = B(x) \log u \qquad \ldots(2.13)$$

where $B(x)$ depends upon x and $u \in (0, 1]$.

Again, setting $u = 1$, $q = 1$ in (2.10), we get

$$\phi_x(v, p) = \phi_x(1, p) + \phi_x(v, 1). \qquad \ldots(2.14)$$

Now, Eqns. (2.12), (2.13) and (2.14) together yield

$$\phi_x(v, p) = A(x) \log p + B(x) \log v, \text{ for } v, p \in (0, 1]. \qquad \ldots(2.15)$$

From (2.9), it follows that

$$f(x, ux, px) = A(x) \log p + B(x) \log u + Cx \log x, \qquad \ldots(2.16)$$
$$\text{for } x \in [0, 1], \quad u, p \in (0, 1].$$

or

$$f(x, u, p) = A(x) \log(p/x) + B(x) \log(u/x) + Cx \log x, \qquad \ldots(2.17)$$
$$\text{for } u, p \leqslant x \text{ and } (x, u, p) \in J.$$

From (2.6) and (2.16), we have

$$A(x) + A(1 - x) = A(1) \text{ and } B(x) + B(1 - x) = B(1), \qquad \ldots(2.18)$$
$$\text{for } x \in [0, 1].$$

The only continuous solutions of the functional equations in (2.18) are given by (Aczel 1966)

$$A(x) = A \cdot x \text{ and } B(x) = B \cdot x, \qquad \ldots(2.19)$$

where A and B are arbitrary constants.

Now, (2.17) becomes

$$f(x, u, p) = A \,.\, x \log(p/x) + B \,.\, x \log(u/x) + Cx \log x, \qquad \ldots(2.20)$$

for $u, p \leqslant x$ and $(x, u, p) \in J$.

Allowing $u, p \leqslant x$, $w = 1 - u \geqslant 1 - x$, $t = 1 - p \geqslant 1 - x$ and $v, q \leqslant 1 - x$ in (2.5) and utilizing (2.16), we get

$$\begin{aligned} &A(x) \log(pq/x) + B(x) \log(uv/x) + Cx \log x \\ &\quad + A(1-x) \log\{(1-p)\, q/(1-x)\} + B(1-x) \log\{(1-u)\, v/(1-x)\} \\ &\quad + C(1-x) \log(1-x) \\ &\quad = A(x) \log(p/x) + B(x) \log(u/x) + Cx \log x \\ &\quad + f(1-x, 1-u, 1-p) + A(1) \log q + B(1) \log v \qquad \ldots(2.21) \end{aligned}$$

which, with the help of (2.18) and (2.19), gives

$$\begin{aligned} &f(1-x, 1-u, 1-p) \\ &\quad = A(1-x) \log\{(1-p)/(1-x)\} + B \,.\, (1-x) \log\{(1-u)/(1-x)\} \\ &\quad + C \,.\, (1-x) \log(1-x), \qquad \ldots(2.22) \end{aligned}$$

for $1 - u, 1 - p \geqslant 1 - x$ and $(1 - x, 1 - u, 1 - p) \in J$.

Thus, (2.20) and (2.22) imply

$$f(x, u, p) = A \,.\, x \log(p/x) + B \,.\, x \log(u/x) + C \,.\, x \log x, \qquad \ldots(2.23)$$

for all $(x, u, p) \in J$.

This completes the proof of the theorem.

3. Information Improvement: Characterization

Let $P = (p_1, p_2, \ldots, p_m)$, $\sum_{i=1}^{m} p_i = 1$, $Q = (q_1, q_2, \ldots, q_m)$, $\sum_{i=1}^{m} q_i = 1$ and $R = (r_1, r_2, \ldots, r_m)$, $\sum_{i=1}^{m} r_i = 1$ be three probability distributions associated with a discrete random variate $X = (x_1, x_2, \ldots, x_m)$. The information theoretic measure that can be associated with P, Q and R satisfying the additivity, viz.,

$$I(P_1 * P_2 \,\|\, Q_1 * Q_2 \,\|\, R_1 * R_2) = I(P_1 \,\|\, Q_1 \,\|\, R_1) + I(P_2 \,\|\, Q_2 \,\|\, R_2) \qquad \ldots(3.1)$$

may be put as

$$\begin{aligned} I(P \,\|\, Q \,\|\, R) &= \sum_{i=1}^{m} f(p_i, q_i, r_i) \\ &= A \sum_{i=1}^{m} p_i \log(r_i/p_i) + B \sum_{i=1}^{m} p_i \log(q_i/p_i) \\ &\quad + C \sum_{i=1}^{m} p_i \log p_i, \qquad \ldots(3.2) \end{aligned}$$

where the base of the logarithm is arbitrary (>1) and A, B and C are arbitrary constants. These constants may be determined by requiring this measure to satisfy some reasonable conditions.

The measure $I(P \parallel Q \parallel R)$, given in (3.2), is a linear combination of entropy of P, relative information of P with respect to Q and that of P with respect to R. This may admit some interesting interpretation, as the difference of two relative informations has the interpretation of being information improvement (Theil 1967).

We now obtain Theil's measure from (3.2). This is done in the next theorem by imposing suitable conditions.

Theorem 2—The additive information theoretic measure (3.2) associated with three distributions P, Q and R of a discrete random variate under the conditions:

$$I(\{1,0\} \parallel \{1,0\} \parallel \{\tfrac{1}{2},\tfrac{1}{2}\}) = -1, \quad \ldots(3.3)$$

$$I(\{\tfrac{1}{2},\tfrac{1}{2}\} \parallel \{\tfrac{1}{2},\tfrac{1}{2}\} \parallel \{\tfrac{1}{2},\tfrac{1}{2}\}) = 0, \quad \ldots(3.4)$$

$$I(\{1,0\} \parallel \{\tfrac{1}{2},\tfrac{1}{2}\} \parallel \{\tfrac{1}{2},\tfrac{1}{2}\}) = 0, \quad \ldots(3.5)$$

is Theil's (1967) information improvement, given by

$$I(P \parallel Q \parallel R) = \sum_{i=1}^{n} p_i \log_2 (r_i/q_i). \quad \ldots(3.6)$$

PROOF: (3.2) with the help of (3.3), (3.4) and (3.5) gives $A = 1/\log 2$, $C = 0$ and $B = -1/\log 2$. Substituting the values of these constants in (3.2), we get (3.6).

4. Non-negativity of $I(P \parallel Q \parallel R)$ and Partitioning of Events

The measure $I(P \parallel Q \parallel R)$, in general, can assume positive as well as negative values. However, in practice, since we shall be interested in positive improvements, it is interesting to examine the condition under which $I(P \parallel Q \parallel R)$ is always positive. This is done in this section. The effect of partitions is also explored. The method uses dynamic programming and intermediatary results are given in the following lemmas.

Lemma 1—Let $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$ be positive numbers, such that $\sum_{i=1}^{n} q_i =$ constant. Then

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

$$\geqslant \sum_{i=1}^{n} p_i \log q_i - \left(\sum_{i=1}^{n} p_i\right) \log \left(\sum_{i=1}^{n} q_i\right). \quad \ldots(4.1)$$

PROOF: Let $\sum_{i=1}^{n} q_i = M$. Thus, it is a n-stage process and the quantity M has been divided into n parts. Further, let

$$Z = q_1^{p_1} \cdot q_2^{p_2} \cdots q_n^{p_n}$$

and

$$f_n(M) = \max Z, \text{ such that } \sum_{i=1}^{n} q_i = M,\ q_i > 0.$$

We first choose q_n arbitrarily and the remaining $(n-1)\, q_i$ so as to get the optimum distribution for $(M - q_n)$. Then, by Bellman's principle of optimality (1957) we get the functional equation:

$$f_n(M) = \max_{0 \leqslant q_n \leqslant M} [q_n^{p_n} f_{n-1}(M - q_n)]$$

where

$$f_1(M) = M^{p_1},$$

so that

$$f_2(M) = \max_{0 \leqslant q_2 \leqslant M} [q_2^{p_2} \cdot (M - q_2)^{p_1}]$$

$$= \left(\frac{p_1}{p_1 + p_2}\right)^{p_1} \left(\frac{p_2}{p_1 + p_2}\right)^{p_2} \cdot M^{p_1+p_2},$$

since the maximum occurs when

$$q_2 = \frac{p_2 M}{p_1 + p_2}.$$

Next,

$$f_3(M) = \max_{0 \leqslant q_3 \leqslant M} \left[q_3^{p_3} \left(\frac{p_1}{p_1 + p_2}\right)^{p_1} \left(\frac{p_2}{p_1 + p_2}\right)^{p_2} (M - q_3)^{p_1+p_2} \right]$$

$$= \left(\frac{p_1}{\sum_{i=1}^{3} p_i}\right)^{p_1} \left(\frac{p_2}{\sum_{i=1}^{3} p_i}\right)^{p_2} \left(\frac{p_3}{\sum_{i=1}^{3} p_i}\right)^{p_3} \cdot M^{p_1+p_2+p_3},$$

since the maximum now occurs at

$$q_3 = \frac{p_3 M}{p_1 + p_2 + p_3}.$$

By induction, it can be easily proved that

$$f_n(M) = \left(\frac{p_1}{\sum_{i=1}^{n} p_i}\right)^{p_1} \left(\frac{p_2}{\sum_{i=1}^{n} p_i}\right)^{p_2} \cdots \left(\frac{p_n}{\sum_{i=1}^{n} p_i}\right)^{p_n} \cdot M^{p_1+p_2+\cdots+p_n},$$

Thus, we have

$$\left(\frac{p_1}{\sum_{i=1}^{n} p_i}\right)^{p_1}\left(\frac{p_2}{\sum_{i=1}^{n} p_i}\right)^{p_2}\cdots\left(\frac{p_n}{\sum_{i=1}^{n} p_i}\right)^{p_n}. M^{p_1+p_2+\cdots+p_n}$$

$$\geqslant q_1^{p_1} . q_2^{p_2} \ldots q_n^{p_n},$$

where

$$\sum_{i=1}^{n} q_i = M, q_i > 0$$

or

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

$$\geqslant \sum_{i=1}^{n} p_i \log q_i - (\sum_{i=1}^{n} p_i) \log (\sum_{i=n}^{n} q_i).$$

From this follows the following result:

Lemma 2—If $\sum_{i=1}^{n} q_i = 1$ and there is a set $r_1, r_2 \ldots, r_n$ such that $\sum_{i=1}^{n} r_i = 1$, each r_i being proportional to the corresponding p_i, then

$$\sum_{i=1}^{n} p_i \log r_i \geqslant \sum_{i=1}^{n} p_i \log q_i. \qquad \ldots(4.2)$$

Remarks: The results of Lemmas 1 and 2 generalize the well-known result (Feinstein 1958)

$$\sum_{i=1}^{n} p_i \log p_i \geqslant \sum_{i=1}^{n} p_i \log q_i,$$

when

$$\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i. \qquad \ldots(4.3)$$

From these two lemmas, we can make the following statement about the information improvement.

Theorem 3—Information improvement obtained by revising the distribution Q as R on the basis of P is non-negative if the elements of R are proportional to those of P.

In the next theorem, we shall come to a result on partitioning of events. We first give a result in Lemma 3.

Lemma 3—If $\sum_{i=1}^{n} p_i \leqslant 1$, each t_i being proportional to p_i, then, we have

$$\sum_{i=1}^{n} I(p_i, u_i, t_i) \geqslant I(\sum_{i=1}^{n} p_i, \sum_{i=1}^{n} u_i, \sum_{i=1}^{n} t_i), u_i > 0, \qquad \ldots(4.4)$$

where

$$I(x, y, z) = \begin{cases} x \log (z/y), & x, y, z > 0 \\ 0 \quad , & x = 0, \quad z \geqslant 0, y > 0 \end{cases}$$

PROOF: Setting

$$q_i = \frac{u_i}{\sum_{i=1}^{n} u_i} \quad \text{and} \quad r_i = \frac{t_i}{\sum_{i=1}^{n} t_i} \quad \text{in (4.2)},$$

we have

$$\sum_{i=1}^{n} p_i \log \frac{t_i}{\sum_{i=1}^{n} t_i} \geqslant \sum_{i=1}^{n} p_i \log \frac{u_i}{\sum_{i=1}^{n} u_i}$$

or

$$\sum_{i=1}^{n} p_i \log \frac{t_i}{u_i} \geqslant \sum_{i=1}^{n} p_i \log \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} u_i}, \quad \sum_{i=1}^{n} p_i \leqslant 1$$

or

$$\sum_{i=1}^{n} I(p_i, u_i, t_i) \geqslant I\left(\sum_{i=1}^{n} p_i, \sum_{i=1}^{n} u_i, \sum_{i=1}^{n} t_i\right).$$

Now, we come to a result on partitioning of events. In what follows, if P, Q and R are generalized distributions, then

$$I(P \parallel Q \parallel R) = \sum_{i=1}^{n} p_i \log (r_i/q_i)/W(P), \text{ where } W(P) = \sum_{i=1}^{n} p_i. \qquad \ldots(4.5)$$

Theorem 4—Let the generalized probability distributions $P = (p_1, p_2, \ldots, p_n)$, $Q = (q_1, q_2, \ldots, q_n)$ and $R = (r_1, r_2, \ldots, r_n)$ be partitioned into m-disjoint subsets $P_1, P_2, \ldots, P_m$, $Q_1, Q_2, \ldots, Q_m$ and $R_1, R_2, \ldots, R_m$ respectively of $n_1, n_2, \ldots, n_m$ elements, such that $\sum_{i=1}^{m} n_i = n, P_i = \sum_{k=n_i-1}^{ni} p_{k-1} = W(P_i)$,

or each i with $n_0 = 0$, then

$$\sum_{i=1}^{m} W(P_i) I(P_i \parallel Q_i \parallel R_i)$$
$$\geqslant W(P) I(P_1, \ldots, P_m \parallel Q_1, \ldots, Q_m \parallel R_1, \ldots, R_m), \qquad \ldots(4.6)$$

where

$$W(P) = \sum_{i=1}^{m} W(P_i).$$

PROOF: For proving the theorem, it is enough to establish the result for two partitions of P, Q, R, viz., $P=(P_1, P_2)$, $Q=(Q_1, Q_2)$, $R=(R_1, R_2)$ of the first n_1 and the remaining $n-n_1$ elements.

Now, from Lemma 3, we have

$$\sum_{i=1}^{n_1} I(p_i, q_i, r_i) \geqslant I\left(\sum_{i=1}^{n_1} p_i, \sum_{i=1}^{n_1} q_i, \sum_{i=1}^{n_1} r_i\right) = I(P_{(1)} \parallel Q_{(1)} \parallel R_{(1)}) \quad \ldots(4.7)$$

where

$$P_{(1)} = \sum_{i=1}^{n_1} p_i, \text{ etc.}$$

and

$$\sum_{i=n_1+1}^{n} I(p_i, q_i, r_i) \geqslant I\left(\sum_{i=n_1+1}^{n} p_i, \sum_{i=n_1+1}^{n} q_i, \sum_{i=n_1+1}^{n} r_i\right)$$

$$= I(P_{(2)} \parallel Q_{(2)} \parallel R_{(2)}), \text{ where } P_{(2)} = \sum_{i=n_1+1}^{n} p_i. \quad \ldots(4.8)$$

Using (4.5) and adding (4.7) and (4.8), we get

$$W(P_1) I(P_1 \parallel Q_1 \parallel R_1) + W(P_2) I(P_2 \parallel Q_2 \parallel R_2)$$

$$\geqslant W(P) I(P_{(1)}, P_{(2)} \parallel Q_{(1)}, Q(2) \parallel R_{(1)}, R_{(2)}).$$

Corollary—The information improvement of distributions P, Q and R, which are unions of $P_1, P_2, \ldots, P_m$, $Q_1, Q_2, \ldots, Q_m$ and $R_1, R_2, \ldots, R_m$ respectively is given by

$$I(P \parallel Q \parallel R) = I(P_1 \cup P_2 \cup \ldots \cup P_m \parallel Q_1 \cup Q_2 \cup \ldots \cup Q_m \parallel$$

$$\parallel R_1 \cup R_2 \cup \ldots \cup R_m)$$

$$= \sum_{i=1}^{m} W(P_i) I(P_i \parallel Q_i \parallel R_i)$$

$$\geqslant I(P_1, P_2, \ldots, P_m \parallel Q_1, Q_2, \ldots, Q_m \parallel R_1, R_2, \ldots, R_m).$$

ACKNOWLEDGEMENT

The author wishes to express his sincere thanks to Dr. Bhu Dev Sharma, Reader in Mathematics, University of Delhi, for guidance and encouragement in carrying out this research work and to S. Surjeet Singh; Vice-Principal, S.G.T.B. Khalsa (E) College, for encouragement.

References

Aczel, J. (1966). Lectures on Functional Equations and Their Applications. Academic Press, New York.

Bellman, R. (1957). Dynamic Programming. Princeton University Press, N.J.

Chaundy, T. W., and McLeod, J. B. (1960). On a functional equation. *Proc. Edinb. math. Soc. math. Notes*, **43**, 7–8.

Feinstein, A. (1958). Foundations of Information Theory. McGraw-Hill Book Co., Inc., New York.

Kannappan, P. L. (1972). On Shannon's entropy, directed divergence and inaccuracy. *Z. Wahrs. Verw Geb.*, **22**, 95–100.

Kerridge, D. F. (1961). Inaccuracy and inference. *Jl. R. Statist. Soc. Ser. B*, **23**, 184–94.

Kullback, S. (1959). Information Theory and Statistics. John Wiley and Sons, Inc., New York.

Sharma, B. D., and Ram Autar (1973). Information improvement functions. *Econometrica*.

Shannon, C. E. (1948). A mathematical theory of communications. *Bell Syst. tech. J.*, **27**, 387–426; 623–56.

Theil, H. (1967). Economics and Information Theory. North Holland Publishing Co., N.Y.