KYBERNETIKA — VOLUME 16 (1980), NUMBER 1

On Additive and Non-Additive Measures of Directed Divergence

M. BEHARA, P. NATH

The axiomatic characterization of some additive and non-additive measures of divergence without assuming the prior existence of any parameter or parameters occurring in their mathematical forms are studied.

1. INTRODUCTION

Let $(\Omega, \mathscr{B}, \mu)$ be a probability space, i.e., Ω is an abstract non-empty set, $\mathscr{B}$ a σ-algebra of subsets of Ω and μ a probability measure defined on $\mathscr{B}$. We ask the following question: How do the two events $E \in \mathscr{B}$ and $F \in \mathscr{B}$ differ from each other? The object of this paper is to give a suitable answer to this question. We shall be considering only events which occur with non-zero probabilities.

Let $\mu(E_1) = p \in I_0$, $\mu(E_2) = q \in I_0$, $I_0 = (0, 1]$. We assume that the amount by which E_1 differs from E_2 is measurable quantitatively and is a function of the probabilities with which the events E_1 and E_2 occur.

2. POSTULATES AND THEIR INDEPENDENCE

Let $F\colon (0, 1] \times (0, 1] \to R$ and $F(p, q)$ denotes the amount by which the event E_1 differs from the event E_2, $\mu(E_1) = p \in I_0$, $\mu(E_2) = q \in (0, 1]$. We shall call $F(p,q)$ the amount of directed divergence of E_1 with respect to E_2. Based on intuition, we assume that F satisfies the following postulates.

Postulate 1. The mapping $p \to F(p, 1)$ is continuous, $p \in I_0$.

According to this postulate, we are comparing an event E, occurring with non-zero probability p, with the sure event Ω occurring with probability one. Consequently,

$F(p, 1)$ denotes the amount by which an event occurring with probability $p \in (0, 1]$ differs from the sure event. Postulate 1 states that if each event, occurring with non-zero probability, is compared with the sure event, then a slight change in the probability of occurrence of the event to be compared with the sure event does not cause an abrupt change in the corresponding amount of directed divergence.

Postulate 2. Let E_1, E_2 and E_3 be any three events occurring with non-zero probabilities p, q and r respectively. Then

$$F(p, r) = F(p, q) + F(q, r), \quad p, q, r \in I_0 . \tag{2.1}$$

Postulate 2 states that the amount by which E_1 differs from E_3 is the sum of the amounts by which E_1 differs from E_2 and E_2 differs from E_3.

Postulate 3. The mapping $p \to F(p, 1)$ is additive in the sense that

$$F(pq, 1) = F(p, 1) + F(q, 1), \quad p, q \in I_0 . \tag{2.2}$$

If the events E_1 and E_2 occurring with non-zero probabilities p and q are independent, then, the probability of their simultaneous occurrence is pq. Postulate 3 states that if all the three events E_1, E_2 and $E_1 \cap E_2$ are compared with the sure event, then the amount by which $E_1 \cap E_2$ differs from the sure event is the sum of the amounts by which the events E_1 and E_2 differ separately from the sure event provided E_1 and E_2 are independent. Postulate 3 is an additivity postulate.

It is obvious that Postulate 1 is independent of Postulates 2 and 3. Any mapping F: $(0, 1] \times (0, 1] \to R$ satisfying Postulate 2 need not satisfy Postulate 3. For example, $F(p, q) = p - q$ satisfies (2.1) but $p \to F(p, 1)$, $p \in I_0$, is not additive. Let us put $r = 1$ in (2.1). Then

$$F(p, q) = F(p, 1) - F(q, 1) . \tag{2.3}$$

Thus, if $p \to F(p, 1)$ is not continuous, then F will no longer be continuous. This shows that Postulate 2 is also independent of Postulate 1.

In the theory of functional equations, it is known, cf. [1], that (2.2) has also discontinuous solutions. Hence, Postulate 3 is independent of Postulate 1.

In order to show that Postulate 3 is also independent of Postulate 2, it is important to observe as to how $F(p, q)$ and $F(p, 1)$ are related to each other. For example, suppose $F(p, q) = \Phi(q) \log p$. Then $p \to F(p, 1)$ is additive but F does not satisfy (2.1). On the other hand, if $p \to F(p, 1)$ and F are related by (2.3), then (2.1) will always be satisfied irrespective of the fact whether $p \to F(p, 1)$ satisfies Postulate 3 or not.

The above observations reveal that Postulates 1, 2 and 3 are independent of each other. In addition to these, we also assume the following normalization postulate:

Postulate 4. $F(1, \frac{1}{2}) = 1$.

This postulate states that a sure event differs from an event occurring with probability $\frac{1}{2}$ by one unit.

Now we prove the following theorem:

Theorem 1. Postulates 1, 2, 3 and 4 characterize the directed divergence $F = I_1$ where

$$I_1(p, q) = \log_2 (p/q) . \tag{2.4}$$

Proof. Postulate 2 implies (2.3). Also, Postulates 1 and 3 give $F(p, 1) = c \log_2 p$ where c is an arbitrary constant. Consequently, (2.3) gives $F(p, q) = c \log_2 (p/q)$. Making use of Postulate 4, we get $c = 1$ so that $F(p, q) = \log_2 (p/q)$. This proves Theorem 1.

From (2.1), it is obvious that if we put $q = p$, we get $F(p, p) = 0$. We put this result in the form of a postulate:

Postulate 5. $F(p, p) = 0$ for all $p \in I_0$. (*Nilpotence*)

Thus, it is clear that Postulate 2 implies Postulate 5 but the converse is not necessarily true. For example, take $F = I_\alpha$, $\alpha \neq 1$, where

$$I_\alpha(p, q) = \frac{p^{\alpha-1}q^{1-\alpha} - 1}{2^{\alpha-1} - 1} , \quad \alpha \neq 1 , \quad p, q \in I_0 . \tag{2.5}$$

Obviously, $I_\alpha(p, p) = 0$ for all $p \in (0, 1]$ but I_α does not satisfy (2.1).

Another interesting consequence of Postulate 2 is

$$F(p, 1) = -F(1, p) , \quad p \in I_0 . \tag{2.6}$$

This follows immediately from (2.3) by putting $p = 1$ and using the fact that $F(1, 1) = 0$ which is also a consequence of Postulate 2. Then, from (2.3) and (2.6), it follows that

$$F(p, q) = -F(q, p) , \quad p, q \in I_0 . \tag{2.7}$$

With these observations, we can prove the following theorem:

Theorem 2. If $F: I_0 \times I_0 \to R$ satisfies Postulates 2 and 3, then

$$F(1, pq) = F(1, p) + F(1, q) , \quad p, q \in I_0 . \tag{2.8}$$

$$F(px, qy) = F(p, q) + F(x, y) , \quad p, q, x, y \in I_0 . \tag{2.9}$$

Proof. (2.8) follows immediately from Postulate 3 and equation (2.6). To prove (2.9), we have

$$\begin{aligned} F(px, qy) &\overset{(2.1)}{=} F(px, 1) + F(1, qy) \\ &\overset{(2.2)}{=} F(p, 1) + F(x, 1) + F(1, q) + F(1, y) \\ &= [F(p, 1) + F(1, q)] + [F(x, 1) + F(1, y)] \\ &\overset{(2.6)}{=} [F(p, 1) - F(q, 1)] + [F(x, 1) - F(y, 1)] \\ &\overset{(2.3)}{=} F(p, q) + F(x, y) . \end{aligned}$$

This completes the proof of Theorem 2.

It should be noted that the additivity of $p \to F(p, 1)$ alone does not necessarily imply that $p \to F(1, p)$ will also be additive. In fact, it all depends upon the form of F. For example, consider $F(p, q) = q \log (p/q)$. Then $p \to F(p, 1)$ is additive but $p \to F(1, p)$ is not. Also, even if both $p \to F(p, 1)$ and $p \to F(1, p)$ are separately additive, still it is not necessary that F will be additive in the sense of (2.9). For example, consider the function F defined by

$$F(p, q) = \log p + \log q + (\log p)(\log q), \quad p, q \in I_0 . \tag{2.10}$$

It is easily seen that $p \to F(p, 1)$ and $p \to F(1, p)$ are additive but F is not. These observations reveal the importance of Postulate 2. Also, we should like to mention that, in the theory of functional equations, equation (2.1) is known as Sincov's functional equation (cf. [1], p. 223).

Not every function F satisfying Sincov's equation (2.1) is additive. However, it turns out to be additive if $p \to F(p, 1)$ satisfies (2.2). Equation (2.2) is a particular case of the functional equation

$$F(pq, 1) = \Phi(F(p, 1), F(q, 1)), \quad p \in I_0, \quad q \in I_0 \tag{2.11}$$

where Φ: $R \times R \to R$ is a polynomial of its argument. Note that (2.2) corresponds to the case when $\Phi(u, v) = u + v$, $u \in R$, $v \in R$.

From intuitive point of view, it is natural to assume F to be a non-constant function because if F is assumed to be constant then this would mean that any two events occurring with non-zero probabilities differ by the same amount and this certainly looks unnatural. In view of this, it is desirable to assume Φ to be a non-constant polynomial of its arguments. Following the arguments as on page 59 of [1], it follows that the only forms of Φ, admissible in (2.11). are

$$\Phi(u, v) = u + v + c , \tag{2.12}$$

$$\Phi(u, v) = Auv + Bu + Bv + \frac{B^2 - B}{A} \tag{2.13}$$

where $A \neq 0$, B and C are arbitrary constants. Consequently, we have

$$F(pq, 1) = F(p, 1) + F(q, 1) + C \tag{2.14}$$

$$F(pq, 1) = AF(p, 1)\, F(q, 1) + BF(p, 1) + BF(q, 1) + \frac{B^2 - BA}{A} \tag{2.15}$$

where $A \neq 0$, B and C are arbitrary constants. Before proving the next theorem, let us state (2.11) in the form of a postulate.

Postulate 6. The mapping $p \to F(p, 1)$ satisfies

$$F(pq, 1) = \Phi(F(p, 1), F(q, 1)), \quad p \in I_0, \quad q \in I_0, \tag{2.11}$$

where Φ: $R \times R \to R$ is a non-constant admissible polynomial of its arguments.

Theorem 3. If F: $I_0 \times I_0 \to R$ satisfies Postulates 2 and 6, then $p \to F(1, p)$ either satisfies (2.8) or

$$F(1, pq) = -AF(1, p)\, F(1, q) + F(1, p) + F(1, q), \quad A \neq 0. \tag{2.16}$$

Likewise, F satisfies either (2.9) or

$$\begin{aligned} F(px, qy) = {} & A[F(p, 1)\, F(x, 1) - F(q, 1)\, F(y, 1)] + \\ & + F(p, q) + F(x, y), \quad A \neq 0. \end{aligned} \tag{2.17}$$

Proof. By Postulate 2, $F(1, 1) = 0$. Hence, (2.14) reduces to (2.2), and (2.8) follows immediately from (2.6) and (2.2). Similarly, putting $q = 1$ in (2.15) and making use of $F(1, 1) = 0$, (2.15) gives $B = 1$ so that (2.15) reduces to

$$F(pq, 1) = AF(p, 1)\, F(q, 1) + F(p, 1) + F(q, 1). \tag{2.18}$$

From (2.6) and (2.18), (2.16) follows immediately. The fact that F satisfies (2.9) under (2.14) with $C = 0$ has been proved in Theorem 2. Now, under (2.15) with $B = 1$,

$$\begin{aligned} F(px, qy) & \overset{(2.1)}{=} F(px, 1) + F(1, qy) \\ & \overset{(2.18)}{\underset{(2.16)}{=}} AF(p, 1)\, F(x, 1) + F(p, 1) + F(x, 1) - AF(1, q)\, F(1, y) + \\ & \qquad\qquad F(1, q) + F(1, y) \\ & \overset{(2.6)}{=} A[F(p, 1)\, F(x, 1) - F(q, 1)\, F(y, 1)] + F(p, q) + F(x, y). \end{aligned}$$

This proves Theorem 3.

The importance of Postulate 6 lies in the fact that, if it is assumed along with Postulate 2, then F has non-additive forms also in addition to additive forms. This is evident from Theorem 3 proved above. The actual forms of F will depend upon the type of regularity conditions imposed upon the mapping $p \to F(p, 1)$, $p \in I_0$.

Making use of Theorem 1, p. 61 in [1], the following theorem now follows immediately:

Theorem 4. If F: $I_0 \times I_0 \to R$ satisfies Postulates 1, 2, 4 and 6, then $F = J_\alpha$ where

$$J_\alpha(p, q) = \frac{p^{\alpha-1} - q^{\alpha-1}}{1 - 2^{1-\alpha}}, \quad \alpha \neq 1 \tag{2.18}$$

$$= \log_2 (p/q), \quad \alpha = 1$$

Clearly, for $\alpha \neq 1$, J_α is non-additive.

3. DECOMPOSABLE DIRECTED DIVERGENCE FUNCTIONS

Definition 1. A function f: $I_0 \times I_0 \to R$ is called a decomposable function if it can be written in the form

$$f(x, y) = \Phi_1(y) = \Phi_2(x), \quad x \in I_0, \quad y \in I_0 \tag{3.1}$$

where Φ_1: $I_0 \to R$ and Φ_2: $I_0 \to R$ ($\Phi_1 = \Phi_2$ is permitted).

It is easy to see that every function f: $I_0 \times I_0 \to R$ which satisfies Sincov's functional equation (2.1) is decomposable but not conversely. Hence, the question arises: When does a decomposable function f: $I_0 \times I_0 \to T$ satisfy Sincov's equation (2.1)? The answer to this question is given by the following theorem which can be easily proved:

Theorem 5. A decomposable function f: $I_0 \times I_0 \to R$ satisfies Sincov's equation (2.1) if and only if

$$\Phi_1(x) = \Phi_2(x) \quad \text{for all} \quad x \in I_0 . \tag{3.2}$$

From Theorem 5, it follows that every directed divergence function F which satisfies Postulate 2 is necessarily decomposable and must be of the form

$$F(p, q) = \Phi_1(q) - \Phi_1(p), \quad p \in I_0, \quad q \in I_0 \tag{3.3}$$

for some function $\Phi_1 : I_0 \to R$. For example, looking at (2.3), we may choose $\Phi_1(x) = -F(x, 1)$, $x \in I_0$. Theorem 4 gives us only decomposable measures of directed divergence. It is enough to choose $\Phi_1 = \psi_\alpha$ where

$$\text{(3.4)} \qquad \psi_\alpha(x) = \frac{1 - x^{\alpha - 1}}{1 - 2^{1-\alpha}}, \quad x \in (0, 1], \quad \alpha \neq 1,$$

$$= \log(1/x), \quad x \in (0, 1], \quad \alpha = 1.$$

Then, for all α,

$$\text{(3.5)} \qquad J_\alpha(p, q) = \psi_\alpha(q) - \psi_\alpha(p), \quad p \in I_0, \quad q \in I_0.$$

We would like to mention that the function ψ_α, defined by (3.4), is the information function of order α introduced by M. Behara and P. Nath in [2].

From intuitive point of view, every directed divergence function F must satisfy Postulate 5 but this alone does not quarantee that F will also be decomposable. The reason is that Postulate 5 does not imply Postulate 2.

Theorem 6. If a decomposable directed divergence $F: I_0 \times I_0 \to R$ satisfies Postulate 2 and is additive, then there exists a function $G: I_0 \to R$ which satisfies Cauchy's equation

$$\text{(3.6)} \qquad G(xy) = G(x) + G(y), \quad x \in I_0, \quad y \in I_0,$$

such that

$$F(p, q) = G(q) - G(p), \quad p, q \in I_0.$$

Proof. Suppose F is a decomposable directed divergence function which satisfies Postulate 2. Then, making use of Definition 1 and Theorem 5, there exists a function $\Phi: I_0 \to R$ such that $F(p, q) = \Phi(q) - \Phi(p)$. Consequently

$$\text{(3.7)} \qquad F(px, qy) = \Phi(qy) - \Phi(px), \quad (p, q, x, y \in I_0).$$

Now, F is additive, that is, it satisfies (2.9). Then, (2.9) and (3.7) give

$$\text{(3.8)} \qquad \Phi(qy) = \Phi(px) - \Phi(q) - \Phi(p) + \Phi(y) - \Phi(x).$$

Putting $q = y = 1$, (3.8) gives

$$\text{(3.9)} \qquad \Phi(px) = \Phi(p) + \Phi(x) - \Phi(1).$$

Define $G: I_0 \to R$ as

$$\text{(3.10)} \qquad G(x) = \Phi(x) - \Phi(1), \quad x \in I_0.$$

Then, G satisfies (3.6) and $F(p, q) = G(q) - G(p)$, $p, q \in I_0$.

Theorem 7. For each function $G\colon I_0 \to R$ satisfying (3.6), there exists a directed divergence function $F\colon I_0 \times I_0 \to R$ which is both additive and decomposable.

Proof. Let G be any function satisfying (3.6). Define F as

$$F(x, y) = G(y) - G(x)\,, \quad x \in I_0\,, \quad y \in I_0\,. \tag{3.11}$$

Then, F satisfies (2.1) and, hence, by Theorem 6, it is decomposable. Now $F(px, qy) = G(qy) - G(px) = G(q) + G(y) - G(p) - G(x) = [G(q) - G(p)] + [G(y) - G(x)] = F(p, q) + F(x, y)$ so that F is also additive.

In the theory of functional equations, there do exist functions G which satisfy (3.6). Since, (3.6) has also discontinuous solutions, therefore, even an additive decomposable directed divergence function can be a discontinuous function. But, from information theory point of view, the discontinuous directed divergence functions are of no use. Hence, we must put some sort of regularity condition on G. We state the following theorem:

Theorem 8. If $G\colon I_0 \to R$ satisfies (3.6) and is bounded from one side on a subset $E \subset I_0$ of positive Lebesgue measure, then every decomposable directed divergence function F, defined by (3.11), is additive and is of the form

$$F(p, q) = \lambda \log (q/p)\,, \tag{3.12}$$

where λ is an arbitrary constant.

A directed divergence function $F\colon I_0 \times I_0 \to R$ which is not decomposable in the sense of Definition 1 will be called an *indecomposable directed divergence function.*

Now, we introduce the following postulate:

Postulate 7. The mapping $F\colon I_0 \times I_0 \to R$ satisfies

$$F(px, qy) = \Phi(F(p, q), F(x, y)) \tag{3.13}$$

where $\Phi\colon R \times R \to R$ is a polynomial of its arguments.

It is obvious that Postulate 7 implies 6 but the converse need not be true. There is no sense in assuming Postulate 2 along with Postulate 7 because Postulate 6, which is a particular case of Postulate 7, together with Postulate 2 determine the forms of F as is evident of Theorem 3. But it does make some sense to assume Postulate 5 with Postulate 7. Then, the complication which arises is that (2.6) no longer holds and hence no information concerning the function $p \to F(1, p)$ can be derived from the function $p \to F(p, 1)$. This difficulty can be overcome by assuming the following postulate:

Postulate 8. The mapping $p \to F(1, p)$ is continuous, $p \in I_0$.

Now we can prove the following theorem:

Theorem 9. Postulate 1, 4, 5, 7 and 8 characterize the directed divergence function $F = I_\alpha$ where

$$(3.14) \qquad I_\alpha(p, q) = \frac{p^{\alpha-1}q^{1-\alpha} - 1}{2^{\alpha-1} - 1}, \quad \alpha \neq 1,$$

$$= \log_2 (p/q), \quad \alpha = 1.$$

Proof. From Postulates 4 and 5, it obviously follows that Φ, in (3.13), cannot be a constant function of its arguments. Let us write (3.13) in the form

$$(3.15) \qquad F(px, qy) = F(p, q) \square F(x, y), \quad (p, q, x, y \in (0, 1]).$$

Then, the operation '$\square$' is both commutative and associative. By following the arguments as on page 59 of [1], it follows that Φ is only of the forms (2.12) and (2.13). Consequently, F satisfies either

$$(3.16) \qquad F(px, qy) = F(p, q) + F(x, y) + C$$

or

$$(3.17) \qquad F(px, qy) = AF(p, q)\, F(x, y) + BF(p, q) + BF(x, y) + \frac{B^2 - B}{A}.$$

where $A \neq 0$, B and C are arbitrary constants.

From (3.16) and Postulates 1, 4, 5 and 8, it follows that $F = I_1$ where $I_1(p, q) = \log (p/q)$. From (3.17), making use of the fact that $F(1, 1) = 0$, a consequence of Postulate 5, it follows that

$$(3.18) \qquad F(p, q) = BF(p, q) + \frac{B^2 - B}{A}.$$

Since $F(p, p) = 0$, it follows that either $B = 0$ or $B = 1$. If $B = 0$, then (3.17) gives

$$(3.19) \qquad F(px, qy) = AF(p, q)\, F(x, y), \quad (x, y, p, q \in (0, 1]).$$

Making use of Postulates 1 and 8, the continuous solutions of (3.19) are of the forms

$$(3.20) \qquad F(p, q) = \frac{p^{\lambda_1} q^{\lambda_2}}{A}$$

where λ_1, λ_2 are arbitrary constants. This solution is not admissible because the RHS of (3.20) does not vanish when $p = q$ and this is a contradiction to Postulate 5. If $B = 1$, then

$$(3.21) \qquad F(px, qy) = AF(p, q)\, F(x, y) + F(p, q) + F(x, y).$$

 Define $h: I_0 \to R$ as

$$h(x, y) = AF(x, y) + 1\,, \quad x, y \in I_0\,. \tag{3.22}$$

Then (3.21) reduces to

$$h(px, qy) = h(p, q)\, h(x, y)\,, \quad (p, q, x, y \in I_0) \tag{3.23}$$

from which it follows that

$$h(p, q) = h(p, 1)\, h(1, q)\,. \tag{3.24}$$

But $p \to h(p, 1)$ and $p \to h(1, p)$ satisfy

$$h(px, 1) = h(p, 1)\, h(x, 1)\,, \tag{3.25}$$

$$h(1, qy) = h(1, q)\, h(1, y)\,. \tag{3.26}$$

From Postulates 1, 8, and equations (3.24), (3.25), (3.26), we have $h(p, q) = p^{\delta_1} q^{\delta_2}$ where $\delta_1 \neq 0$, $\delta_2 \neq 0$ are arbitrary constants. Consequently,

$$F(p, q) = \frac{p^{\delta_1} q^{\delta_2} - 1}{A}\,, \quad A \neq 0\,, \quad \delta_1 \neq 0\,, \quad \delta_2 \neq 0\,. \tag{3.27}$$

Choose $p_0 \in (0, 1)$ arbitrarily. By Postulate 5, $F(p_0, p_0) = 0$. Then, (3.27) gives $\delta_2 = -\delta_1$ so that

$$F(p, q) = \frac{(p/q)^{\delta_1} - 1}{A}\,, \quad \delta_1 \neq 0\,. \tag{3.28}$$

Making use of Postulate 4, we get $A - 2^{\delta_1} - 1$. Choosing $\delta_1 = \alpha - 1$, $\alpha \neq 1$, we get $F = I_\alpha$, $\alpha \neq 1$, where

$$I_\alpha(p, q) = \frac{p^{\alpha - 1} q^{1 - \alpha} - 1}{2^{\alpha - 1} - 1}\,, \quad \alpha \neq 1\,, \quad p, q \in I_0\,. \tag{3.29}$$

This completes the proof of Theorem 9.

It is obvious that I_α, for $\alpha \neq 1$, does not satisfy Sincov's functional equation (2.1). Also, for $\alpha \neq 1$, I_α is an indecomposable directed divergence function.

4. SOME MEASURES OF DIRECTED DIVERGENCE FOR TWO GENERALIZED DISCRETE PROBABILITY DISTRIBUTIONS

Let $\Gamma_m = \{(p_1, p_2, \ldots, p_n) : p_k > 0, \quad k = 1, 2, \ldots, n, \quad \sum_{k=1}^{n} p_k \leqq 1\}$, $n = 1, 2, \ldots$ denote the set of all n-components discrete generalized probability distributions. Let $(p_1, p_2, \ldots, p_n) = P \in \Gamma_n$ and $(q_1, q_2, \ldots, q_n) = Q \in \Gamma_n$. We define the directed divergence $\mathscr{D}^F(p \parallel Q)$ of P with respect to Q as

(4.1) $$\mathscr{D}^F(P \parallel Q) = \sum_{k=1}^{n} p_k F(p_k, q_k) / \sum_{k=1}^{n} p_k$$

where $F: I_0 \times I_0 \to R$ is a directed divergence function.

It is clear that the form of $\mathscr{D}^F(P \parallel Q)$ depends upon the form of F. If $F = J_\alpha$ given by (2.18), then $\mathscr{D}^F(P \parallel Q) = D_\alpha(P \parallel Q)$ where

(4.2) $$D_\alpha(P \parallel Q) = \frac{\sum_{k=1}^{n} p_k^\alpha - \sum_{k=1}^{n} p_k q_k^{\alpha-1}}{\left(\sum_{k=1}^{n} p_k\right)(1 - 2^{1-\alpha})}, \quad \alpha \neq 1$$

$$= \sum_{k=1}^{n} p_k \log (p_k/q_k) / \sum_{k=1}^{n} p_k, \quad \alpha = 1.$$

P. Nath [7] proposed a non-additive measure $h_\alpha(P \parallel Q)$ of inaccuracy

(4.3) $$h_\alpha(P \parallel Q) = \frac{1 - \left(\sum_{k=1}^{n} p_k q_k^{\alpha-1} / \sum_{k=1}^{n} p_k\right)}{1 - 2^{1-\alpha}}, \quad \alpha \neq 1$$

which reduces to non-additive entropy (I. Vajda [11])

$$h_\alpha(P) = \frac{1 - \left(\sum_{k=1}^{n} p_k^\alpha / \sum_{k=1}^{n} p_k\right)}{1 - 2^{1-\alpha}}, \quad \alpha \neq 1$$

when $P \equiv Q$. As $\alpha \to 1$, it can be easily seen that $\lim_{\alpha \to 1} h_\alpha(P \parallel Q) = H_1(P \parallel Q)$ and $\lim_{\alpha \to 1} h_\alpha(P) = H_1(P)$ where

(4.4) $$H_1(P \parallel Q) = \sum_{k=1}^{n} p_k \log q_k / \sum_{k=1}^{n} p_k$$

(4.5) $$H_1(P) = \sum_{k=1}^{n} p_k \log p_k / \sum_{k=1}^{n} p_k .$$

For axiomatic characterizations of $H_1(P \parallel Q)$ and $H_1(P)$, see P. Nath [7] and A. Rényi [9]. Now, it is clear that

(4.6) $$D_\alpha(P \parallel Q) = h_\alpha(P \parallel Q) - h_\alpha(P), \quad \alpha \neq 1$$

$$= H_1(P \parallel Q) - H_1(P), \quad \alpha = 1.$$

If $F = I_\alpha$ given by (3.14), then $\mathscr{D}^F(P \parallel Q) = D_\alpha^*(P \parallel Q)$ where

$$(4.7) \qquad D_\alpha^*(P \parallel Q) = \frac{1 - (\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha} / \sum_{k=1}^{n} p_k)}{1 - 2^{\alpha-1}}, \quad \alpha \neq 1$$

$$= \sum_{k=1}^{n} p_k \log (p_k/q_k) / \sum_{k=1}^{n} p_k, \quad \alpha = 1$$

The measure of directed divergence $D_\alpha^*(P \parallel Q)$, for $\alpha \neq 1$, si due to the second author (cf. [7], [8]). It should be noted that, for $\alpha \neq 1$, $D_\alpha^*(P \parallel Q)$ is non-additive and the same is also true of $D_\alpha(P \parallel Q)$, $\alpha \neq 1$. However, $D_1(P \parallel Q)$ or equivalently $D_\alpha^*(P \parallel Q)$ is additive.

We would like to emphasize that all the measures of directed divergence characterized axiomatically in this paper, *do not assume the prior existence of the parameter α occurring in them.* Any axiomatic characterization involving the parameter α explicitly in the postulates is undesirable because it will make the definition of corresponding measure of directed divergence an artificial one.

(Received May 30, 1977.)

REFERENCES

[1] J. Aczél: Lectures on Functional Equations and Their Applications. Academic Press, New York 1966.

[2] M. Behara, P. Nath: Information and Entropy of Countable Measurable Positions I. Kybernetika *10* (1974), 6, 491–503.

[3] Z. Daróczy: Generalized Information Functions. Information and Control *16* (1970), 1, 36–51.

[4] J. Havrda, F. Charvát: Quantification Method of Classification Processes. The Concept of Structural *a*-entropy. Kybernetika *3* (1967), 1, 30–35.

[5] D. F. Kerridge: Inaccuracy and Inference. J. Roy. Statist. Soc. Ser. B *23* (1961), 184–194.

[6] S. Kullback: Information Theory and Statistics. John Wiley & Sons, New York; Chapman & Hall, London 1959.

[7] P. Nath: An Axiomatic Characterization of Inaccuracy for Discrete Generalized Probability Distributions. Opsearch *7* (1970), 115–133.

[8] P. Nath: Some Axiomatic Characterizations of a Non-Additive Measure of Divergence in Information. Jour. Math. Sciences *7* (1972), 57–68.

[9] A. Rényi: On Measures of Entropy and Information. Proc. 4th Berkeley Symp. Math. Statist. Probability 1 (1960), 547–561.

[10] C. E. Shannon: A Mathematical Theory of Communication. Bell System Technical Journal *27* (1968), 379–423, 623–656.

[11] I. Vajda: Axioms for *a*-entropy of a Generalized Probability Scheme (Czech). Kybernetika *4* (1968), 2, 105–112.

Professor Dr. M. Behara, Dr. P. Nath, Department of Mathematics, McMaster University, Hamilton, Ontario, Canada.