SELF-INFORMATION AND INFORMATION

by Samir Kumar Mukherjee, *Department of Mathematics, Krishna Chandra College, P.O. Hetampur Rajbati, Dist. Birbhum, West Bengal*

(*Received* 1 *July* 1976)

The paper starts with a triparametric self-information function and correspondingly, a triparametric entropy. Some familiar entropies are derived as particular cases. A new measure called 'information deviation' and some generalizations of Kullback information are obtained under changed boundary conditions.

1. Introduction

Shannon (1948) first introduced the idea of self-information function in the form

$$f(x) = -\log_2 x, \ (0 < x \leqslant 1). \qquad \text{...(1)}$$

The method of averaging self-informations introduced by Shannon is readopted by us in this paper. Like Shannon, we introduce a triparametric self-information function defined by

$$f_3(x; \alpha, \beta, \gamma) = C(x^\alpha - x^\beta)/x^\gamma, \quad (0 < x \leqslant 1, \ \alpha \geqslant 0, \ \beta \geqslant 0, \ \nu \geqslant 0, \ \alpha \neq \beta) \qquad \text{...(2)}$$

where C is a constant, depending upon the real-valued parameters α, β, γ and C is ascertained by a suitable pair (x, f_3). We apply the following conditions on f_3:

(i) $f_3 \to \infty$ as $x \to 0$, when $\gamma >$ at least one of α, β;

(ii) $f_3 = 0$, when $x = 1$; or, $f_3 \to 0$, when $x \to 0$ for $\alpha, \beta > \gamma$.

(iii) $f_3 = 1$, when $x = \frac{1}{2}$. Then $C = (2^{\gamma-\alpha} - 2^{\gamma-\beta})^{-1}$.

The function shows the following particular behaviours:

I. If α, β are fixed, then for $x < \frac{1}{2}$, $f_3 \to \infty$ as $\gamma \to \infty$; and for $x > \frac{1}{2}$, $f_3 \to 0$, as $\gamma \to \infty$.

II. For any fixed γ, $f_3 \to -(2x)^{\beta-\gamma} \log_2 x$, as $\alpha \to \beta$.

III. If $\beta = \gamma$ and $\alpha (< \gamma) \to \gamma$, then $f_3 \to -\log_2 x$.

Self-information function is different from information function. Different authors, namely Daroczy (1970), Aczel (1966), Chaundy and Mcleod (1960), Havrda and Charvat (1967), Kannapan (1972), Sharma and Taneja (1975), Mittal (1975) and some others have solved some typical functional equations and have used their solutions as entropy, inaccuracy, directed divergence, etc., in the capacity of finite measures only in complete probability distributions. The method of averaging self-informations includes the case of generalized probability distributions. Moreover, we have discussed in this paper, information measures in the capacity of even an infinite range, because a parameter can have negative values also corresponding to phenomenal circumstances. Further, since it is uncertain and difficult to choose an arbitrary functional equation and to find its suitable solutions to be used as information measures, it becomes easier, if we choose any suitable parametric self-information function that can satisfy a number of effective boundary conditions. We have given a most simple and general choice in (2), which is the basis of the present paper.

Section 2 describes a triparametric entropy from which other familiar entropies have been deduced as particular cases. We have given a number of applications of this entropy in section 3 as joint entropy, triparametric information function, generalized information function, generalized inaccuracy, a new information called 'information deviation' and lastly generalizations of Kullback's information.

2. Triparametric Entropy

Let $P = (p_1, p_2, \ldots, p_n)$ be a finite discrete probability distribution, where

$$0 < p_k \leqslant 1, \; w(P) = \sum_1^n p_k \leqslant 1.$$

Then, averaging the function $f_3(p_k; \alpha, \beta, \gamma)$ with respect to P, we define the triparametric entropy as

$$H_3(P; \alpha, \beta, \gamma) = (2^{\gamma-\alpha} - 2^{\gamma-\beta})^{-1} \left[\sum_1^n p_k^{\alpha-\gamma+1} - \sum_1^n p_k^{\beta-\gamma+1} \right] / w(P), \quad \ldots(3)$$

in which the powers of p_k may be positive, zero or negative according to the behaviours of α, β, γ where $\alpha \geqslant 0$, $\beta \geqslant 0$, $\gamma \geqslant 0$, $\alpha \neq \beta$. The suffix i in H_i will indicate the number of parameters used in the entropy function. Writing P^* for P, when P is complete, we have

$$H_3(P^*; \alpha, \beta, \gamma) = (2^{\gamma-\alpha} - 2^{\gamma-\beta})^{-1} \left[\sum_1^n p_k^{\alpha-\gamma+1} - \sum_1^n p_k^{\beta-\gamma+1} \right], \quad \ldots(4)$$

where α, β, γ are as in (3).

2.1. *Some Familiar Entropies*

From (4), we get the following entropies as particular cases:

(i) $\gamma = 1$ gives Sharma and Taneja's entropy (1975) of type (α, β) in the form

$$H_2(P^*; \alpha, \beta) \equiv H_{(\alpha,\beta)}(P) = (2^{1-\alpha} - 2^{1-\beta})^{-1} \left[\sum_1^n p_k^{\alpha} - \sum_1^n p_k^{\beta} \right], \; (\alpha \neq \beta); \quad \ldots(5)$$

and,

$$\underset{\alpha\to\beta}{\text{Lt}}\ H_2(P^*;\alpha,\beta)=\Big(\sum_1^n p_k^\beta \log_2\frac{1}{p_k}\Big)2^{\beta-1}. \qquad \ldots(5')$$

(ii) Putting $\alpha=\gamma=1$, we get Daroczy's entropy (1970) of type β as

$$H_1(P^*;\beta)\equiv H_n^\beta(p_1,p_2,\ldots p_n)=(2^{1-\beta}-1)^{-1}\,[\,\sum_1^n p_k^\beta-1],\ (\beta>0,\ \beta\neq 1) \qquad \ldots(6)$$

and also, Havrda and Charvat's entropy (1967) in the form (6) with $\beta\neq 1$.

(iii) When $\beta=\gamma$ and $\alpha\,(<\gamma)\to\gamma$, then (4) reduces to

$$H_0(P^*)\equiv H(p_1,p_2,\ldots,p_n)=\sum_1^n p_k\log_2\frac{1}{p_k}, \qquad \ldots(7)$$

which is Shannon's entropy (1948).

(iv) Another uniparametric entropy is obtained by putting $\gamma=\beta=2\alpha$ in the form

$$H_1{}^*(p^*;\alpha)=(2^\alpha-1)^{-1}\,[\,\sum_1^n p_k^{1-\alpha}-1],\quad (\alpha>0), \qquad \ldots(8)$$

whose characteristics, properties and comparisons with other entropies have been reported by the author (Mukherjee) elsewhere.

(v) When $n>2$, then $H_3\to\infty$ as $\gamma\to\infty$ for any fixed α, $\beta<\gamma$; when $n=1$, then $H_3=0$, $p_1=1$; and when $n=2$, then $H_3=1$ or $H_3\to\infty$ as $\gamma\to\infty$.

3. Applications of the Entropy (4)

3.1. *Joint Entropy*

We assert that for joint probability distribution, a relation analogous to Eqn. (4) also holds in the form

$$H_3[(PQ)^*;\alpha,\beta,\gamma]=(2^{\gamma-\alpha}-2^{\gamma-\beta})^{-1}\,[\,\sum_{k=1}^n\sum_{j=1}^m p_{kj}^{\alpha-\gamma+1}-\sum_{k=1}^n\sum_{j=1}^m p_{kj}^{\beta-\gamma+1}],$$

$$(0<p_{kj}\leqslant 1,\ \sum_{k=1}^n\sum_{j=1}^m p_{kj}=1;\ \alpha\geqslant 0,\ \beta\geqslant 0,\ \gamma\geqslant 0,\ \alpha\neq\beta). \qquad \ldots(9)$$

Theorem 1—If $P^*=(p_1,p_2,\ldots,p_n)$ be the distribution of input symbols of a source, $Q^*=(q_1,q_2,\ldots,q_m)$ be that of output symbols, and $(PQ)^*=(p_{k1},p_{k2},\ldots,p_{km};\ k=1,2,\ldots,n)$ be the joint distribution of input and output symbols; also,

$$R_k^*=\Big(\frac{p_{k1}}{p_k},\ \frac{p_{k2}}{p_k},\ldots,\ \frac{p_{km}}{p_k}\Big)$$

be the conditional distribution of output symbols and

$$R_j^* = \left(\frac{p_{1j}}{q_j}, \frac{p_{2j}}{q_j}, \ldots, \frac{p_{nj}}{q_j}\right)$$

be the conditional distribution of input symbols, where,

$$p_{kj}/p_k = p_{j|k}, (j = 1, 2, \ldots, m); \; p_{kj}/q_j = p_{k|j}, \; (k = 1, 2, \ldots, n);$$

$$\sum_{j=1}^{m} p_{kj} = p_k \quad \text{and} \quad \sum_{k=1}^{n} p_{kj} = q_j,$$

then

$$H_3 [(PQ)^*; \alpha, \beta, \gamma]$$

$$= \sum_{k=1}^{n} p_k^{\beta-\gamma+1} H_3 (R_k^*; \alpha, \beta, \gamma) + (2^{\gamma-\alpha} - 2^{\gamma-\beta})^{-1} \left[\sum_{k=1}^{n} \{(p_k^{\alpha-\gamma+1} - p_k^{\beta-\gamma+1}) \sum_{j=1}^{m} p_{j|k}^{\alpha-\gamma+1}\}\right] \qquad \ldots(10)$$

Putting $\alpha = \gamma = 1$ and using $\sum_{j=1}^{m} p_{j|k} = 1$ in (10), we have

$$H_1 [(PQ)^*; \beta] = \sum_{k=1}^{n} p_k^{\beta} H_1 (R_k^*; \beta) + H_1 (P^*; \beta). \qquad \ldots(11)$$

Theorem 2—If $p_{kj} = p_k q_j$, then

$$H_3 [(PQ)^*; \alpha, \beta, \gamma]$$

$$= \sum_{k=1}^{n} p_k^{\alpha-\gamma+1} H_3 (R_k^*; \alpha, \beta, \gamma) + \sum_{j=1}^{m} q_j^{\beta-\gamma+1} H_3 (R_j^*; \alpha, \beta, \gamma)$$

$$= \sum_{k=1}^{n} p_k^{\alpha-\gamma+1} H_3 (Q^*; \alpha, \beta, \gamma) + \sum_{j=1}^{m} q_j^{\beta-\gamma+1} H_3 (P^*; \alpha, \beta, \gamma). \qquad \ldots(12)$$

3.2. *Triparametric Information Function*

Eqn. (4) helps us to define a triparametric information function in the form

$$F_3 (x) = F_3 (x; \alpha, \beta, \gamma) = (2^{\gamma-\alpha} - 2^{\gamma-\beta})^{-1} (x^{\alpha-\gamma+1} - x^{\beta-\gamma+1}),$$

$$(\alpha \geqslant 0, \beta \geqslant 0, \gamma \geqslant 0, \alpha \neq \beta \text{ and } 0 < x \leqslant 1), \qquad \ldots(13)$$

where $F_3 (0) = 0$ not always, but $F_3 (1) = 0$ and $F_3 (\frac{1}{2}) = \frac{1}{2}$ always.

Thus,

$$H_3 (P^*; \alpha, \beta, \gamma) = \sum_{k=1}^{n} F_3 (p_k), \quad (0 < p_k \leqslant 1, \; \sum_{1}^{n} p_k = 1). \qquad \ldots(14)$$

Putting $a = \alpha - \gamma + 1$, $b = \beta - \gamma + 1$ in (13), we have

$$F_2 (x) \equiv F_2 (x; a, b) = (2^{1-a} - 2^{1-b})^{-1} (x^a - x^b), \quad (-\infty < a, b < \infty, a \neq b). \qquad \ldots(15)$$

Now, from practical point of view, as far as an inaccuracy in a measure is concerned, a measure is associated with at least two probability distributions, corresponding to which at least two variables u and v are needed. This suggests the choice of at least four parameters a, b, c and d.

3.3. *Generalized Information Function*

Concerning an association of two variables u, v and four parameters, a,b, c, d, an information measure analogous to (15) is introduced by

$$F_4(u,v) \equiv F_4(u,v;a,b,c,d)$$
$$= G[u^a v^b - u^c v^d], \quad (0 < u \leqslant 1, 0 < v \leqslant 1, -\infty < a,b,c,d < \infty, \; a \neq b \neq c \neq d), \qquad ...(16)$$

as the generalized information function, which possesses the characteristic of becoming both bounded and unbounded.

Everything will now depend upon one's choice of boundary conditions and parametric values, and the result will vary, if the choices vary. We give below only two sets of boundary conditions and therein one is to decide which is the better of the two.

3.3.1. *Boundary Conditions*

(i) At $u = 1$, $v = \frac{1}{2}$, $F_4(1, \frac{1}{2}) = \frac{1}{2}$, so that $G = (2^{1-b} - 2^{1-d})^{-1}$, where $b \neq d$. If $a + b = c + d$, where $a \neq c$, then $F_4(\frac{1}{2}, \frac{1}{2}) = 0$. Similarly at $u = \frac{1}{2}, v = 1$, $F_4(\frac{1}{2}, 1) = \frac{1}{2}$, so that $G = (2^{1-a} - 2^{1-c})^{-1}$, where $a \neq c$.

(ii) At $u = 1$, $v = \frac{1}{2}$, $F_4(1, \frac{1}{2}) = 1$, so that $G = (2^{-b} - 2^{-d})^{-1}$, where $b \neq d$. At $u = \frac{1}{2}$, $v = 1$, $F_4(\frac{1}{2}, 1) = 1$, so that $G = (2^{-a} - 2^{-c})^{-1}$, where $a \neq c$.

3.3.2. *Generalized Inaccuracy*

Let $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$ be two discrete probability distributions concerned with (16), where $0 < p_k \leqslant 1$, $0 < q_k \leqslant 1$, $\sum_1^n p_k = 1$; $\sum_1^n q_k = 1$, $(u, v) = (p_k, q_k)$ or (q_k, p_k); $k = 1, 2, \ldots, n$.

We may then define the generalized inaccuracies by

$$I_4(P \,\|\, Q) = \sum_1^n F_4(p_k, q_k) = (2^{1-b} - 2^{1-d})^{-1} \left[\sum_1^n p_k^a q_k^b - \sum_1^n p_k^c q_k^d \right], \quad (b \neq d), \qquad ...(17)$$

$$I_4(Q \,\|\, P) = \sum_1^n F_4(q_k, p_k) = (2^{1-b} - 2^{1-d})^{-1} \left[\sum_1^n q_k^a p_k^b - \sum_1^n q_k^c p_k^d \right], \quad (b \neq d), \qquad ...(18)$$

which follow from (16) and 3.3.1 (i).

Given P and Q, we see that

(i) $I_4(P \| Q) \to +\infty$ or $-\infty$, according as $a \to -\infty$ or $c \to -\infty$ for $b < d$; or as $c \to -\infty$ or $a \to -\infty$ for $b > d$.

(ii) If $d = 1$, $c = 0$, then $I_4(P \| Q) \to (1 - 2^{1-b})^{-1}$ as $a \to \infty$.

(iii) If $d = 1$, $c = 0$, then $I_4(P \| Q) \to 1$ as $b \to \infty$.

It is to be noted that when $d = 1$, $c = 0$, then

$$I_2(P \| Q) = (2^{1-b} - 1)^{-1} [\sum_1^n p_k^a q_b^k - 1]. \quad \text{...(19)}$$

3.3.3. *Information Deviations*

If $d = 1$, $c = 0$, $a + b = 1$, then we introduce the quantities

$$D(Q \| P \| Q) = \operatorname*{Lt}_{b \to 1} I_4(P \| Q) = H(Q) - H(Q \| P) \quad \text{...(20)}$$

and

$$D(P \| Q \| P) = \operatorname*{Lt}_{b \to 1} I_4(Q \| P) = H(P) - H(P \| Q) \quad \text{...(21)}$$

as the information deviations of Q from P and of P from Q respectively, where

$$H(P) = \sum_1^n p_k \log_2 \frac{1}{p_k}, \qquad H(Q) = \sum_1^n q_k \log_2 \frac{1}{q_k}$$

are Shannon's entropies and

$$H(Q \| P) = \sum_1^n q_k \log_2 \frac{1}{p_k}, \qquad H(P \| Q) = \sum_1^n p_k \log_2 \frac{1}{q_k}$$

are Kerridge's inaccuracies (1961). Thus,

$$D(Q \| P \| Q) = \sum_1^n q_k \log_2 \frac{p_k}{q_k}, \qquad D(P \| Q \| P) = \sum_1^n p_k \log_2 \frac{q_k}{p_k}. \quad \text{...(22)}$$

Similar definitions of 'information deviations' of order α have been given by the author elsewhere (Mukherjee 1976).

3.3.4. *Kullback's Information and Its Generalizations*

If we take the boundary conditions 3.3.1 (ii), then,

$$I_4(P \| Q) = \tfrac{1}{2} I_4^*(P \| Q), \quad \text{...(23)}$$

where

$$I_4^*(P \parallel Q) = (2^{-b} - 2^{-d})^{-1} [\sum_1^n p_k^a q_k^b - \sum_1^n p_k^c q_k^d], \quad (b \neq d) \qquad \ldots(24)$$

Now, if $d = 0$, $c = 1$, $a + b = 1$, then

$$\underset{b \to 0}{\text{Lt}}\ I_4(P \parallel Q) = \tfrac{1}{2} I(P \parallel P \parallel Q), \underset{b \to 0}{\text{Lt}}\ I_4(Q \parallel P) = \tfrac{1}{2} I(Q \parallel Q \parallel P) \quad \ldots(25)$$

where

$$I(P \parallel P \parallel Q) = \sum_1^n p_k \log_2 \frac{p_k}{q_k} = H(P \parallel Q) - H(P) \qquad \ldots(26)$$

and

$$I(Q \parallel Q \parallel P) = \sum_1^n q_k \log_2 \frac{q_k}{p_k} = H(Q \parallel P) - H(Q) \qquad \ldots(27)$$

represent Kullback's informations.

Information deviations and Kullback's informations are equal and opposite measures. The fact follows from

$$D(Q \parallel P \parallel Q) + I(Q \parallel Q \parallel P) = 0, \; D(P \parallel Q \parallel P) + I(P \parallel P \parallel Q) = 0 \qquad (28)$$

It may be noted that information deviations and Kullback's informations become zero, if $p_k = q_k$ for $k = 1, 2, \ldots, n$.

We shall now show that so far as our generalized inaccuracies (17) and (18) are concerned, there exist certain boundary conditions for which certain limiting functions of (17) or (18) may be taken as the generalized forms of Kullback's informations. For this, we generalize the boundary conditions of 3.3.1 in the following ways and get the results:

3.3.4.1. Let

$$u = 1, \; v = \frac{1}{2}, \; F_4\left(1, \frac{1}{2}\right) = \frac{1}{2^m},$$

where m is a real number $\geqslant 0$. Then, we have for $d = 0$, $c = 1$, $a + b = 1$,

$$I^{(1)}(P, Q; m) = \underset{b \to 0}{\text{Lt}}\ I_4(P \parallel Q) = 2^{-m} \sum_1^n p_k \log_2 \frac{p_k}{q_k}, \quad (m \geqslant 0) \qquad \ldots(29)$$

to be called the first generalized Kullback information. For $m = 0$ in (29), we get Kullback's information. The information (29) decreases as m increases.

3.3.4.2. Let

$$u = 1,\ v = \frac{1}{2},\ F_4\left(1, \frac{1}{2}\right) = \frac{1}{2^m},$$

where m is a real number $\geqslant 0$. Also, let $d = 0$, $c = 1 + m$, and $a + b = 1 + m$. Then, we have

$$I^{(2)}(P, Q; m) = \underset{b \to 0}{\text{Lt}}\ I_4(P \parallel Q) = 2^{-m} \sum_{1}^{n} p_k^{m+1} \log_2 \frac{p_k}{q_k},\ (m \geqslant 0) \quad \ldots(30)$$

to be called the second generalized Kullback information. It is to be observed that $I^{(2)}(P, Q; m) \leqslant I^{(1)}(P, Q; m)$.

For $m = 0$ in (30), we get Kullback's information.

3.3.4.3. Let $u = 1$, $v = \frac{1}{2}$, $F_4(1, \frac{1}{2}) = 2^{-1/m}$, where m is any positive real number. Then, the values $d = 0$, $c = 1 + 1/m$ and $a + b = 1 + 1/m$ lead to the information

$$I^{(3)}(P, Q; m) = \underset{b \to 0}{\text{Lt}}\ I_4(P \parallel Q) = 2^{-1/m} \sum_{1}^{n} p_k^{1/m+1} \log_2 \frac{p_k}{q_k} \quad \ldots(31)$$

which may be called the third generalized Kullback information. In this case

$$\underset{m \to 0}{\text{Lt}}\ I^{(3)}(P, Q; m) = 0 \quad \text{and} \quad \underset{m \to \infty}{\text{Lt}}\ I^{(3)}(P, Q; m) = I(P \parallel P \parallel Q). \quad (32)$$

ACKNOWLEDGEMENT

I am thankful to Dr. Bhudev Sharma, Reader, Department of Mathematics, Delhi University and Dr. Ram Autar, Department of Education in Science and Mathematics, NCERT, New Delhi, for their kind help in the preparation of this paper.

REFERENCES

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

Chaundy, T. W., and Mcleod, J. B. (1960). On a functional equation. *Proc. Edinb. math. Soc. Edinb. math. Notes*, No. 43, 7–8.

Daroczy, Z. (1970). Generalized information functions. *Inf. Control*, **16**, 35–51.

Havrda, J., and Charvat, F. (1967). Quantification method of classification processes. The concept of structural α-entropy. *Kybernetika*, **3**, 30–35.

Kannapan, Pl. (1972). On Shannon's entropy, directed divergence and inaccuracy. *Z. Wahrs. Verw. Ceb.*, **22**, 95–100.

Kerridge, D. F. (1961). Inaccuracy and inference. *J. R. statist. Soc.*, **B23**, 184–94.

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

Mittal, D. P. (1975). On some functional equations concerning entropy, directed divergence and inaccuracy. *Metrika*, **22**, 35–45.

Mukherjee, S. K. (1976). Source-character of information based on the generalized order relations of conjugacy. *Maths. Educ. India*, **10** (1), 17–25.

———— (1976). Some parametric entropies of generalized probability distributions. *Math. Stud.* (accepted).

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. tech. J.*, **27**, 379-423, 623–58.

Sharma, B. D., and Taneja, I. J. (1975). Entropy of type (α, β) and other generalized measures in information theory. *Metrika*, **B 22**, 205–15.