

A FUNCTIONAL EQUATION OF TWO VARIABLES IN INFORMATION THEORY

by BHU DEV SHARMA and RAM AUTAR,* *Faculty of Mathematics, University of Delhi, Delhi 110007*

(Communicated by F.C. Auluck, F.N.A.)

(Received 26 February 1974; after revision 30 August 1974)

In this paper we have formed a functional equation in two variables whose solutions are defined as inaccuracy functions under suitable boundary conditions and measure of inaccuracy is defined in terms of inaccuracy functions. A characterization of Kerridge's (1961) inaccuracy function, under suitable conditions, is given. An inaccuracy function is a generalization of information function studied by Kendall (1964) and Daroczy (1970).

1. INTRODUCTION

According to Kendall (1964), a real valued function $f(x)$ of a real variable $x \in [0, 1]$ is called an information function if it satisfies the functional equation.

$$f(x) + (1-x)f\left(\frac{y}{1-x}\right) = f(y) + (1-y)f\left(\frac{x}{1-y}\right) \quad (1.1)$$

for all $x, y \in [0, 1]$, $x+y \leq 1$ and the boundary conditions

$$f(0) = f(1) ; f\left(\frac{1}{2}\right) = 1. \quad (1.2)$$

For a probability distribution $P = (p_1, \dots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$, the entropy is defined as

$$H_n^f(P) = \sum_{i=1}^n s_i f(p_i/s_i) \quad (1.3)$$

where $s_i = p_1 + p_2 + \dots + p_i$ and f is an information function. If f is Shannon's (1948) information function (Kendall 1964 ; Tverberg 1958 ; Aczel and Daroczy 1975) given by

$$f(x) = \begin{cases} -x \log x - (1-x) \log (1-x) & \text{if } x \in (0,1) \\ 0 & \text{if } x = 0 \text{ or } x = 1 \end{cases} \quad (1.4)$$

then, we have

$$H_n^f(P) = - \sum_{i=1}^n p_i \log p_i \quad (1.5)$$

* Present address : Department of Education in Science and Mathematics, NCERT, Sri Aurobindo Marg, New Delhi-110016

which is Shannon's entropy of probability distribution P .

A generalization of Shannon's idea is given by Kerridge (1961). The measure of inaccuracy suggested by Kerridge has many useful applications in Statistics and in the development of personal codes.

Let X be a discrete random variate taking n finite values (x_1, x_2, \dots, x_n) with probabilities $P = (p_1, p_2, \dots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$. Suppose that an experimenter asserts that the probabilities of happening of the events are $Q = (q_1, q_2, \dots, q_n)$ $q_i \geq 0$, $\sum_{i=1}^n q_i = 1$, then the measure of inaccuracy of Q with respect to P as given by Kerridge is

$$H_n(P; Q) = - \sum_{i=1}^n p_i \log q_i \quad (1.6)$$

If $q_i = 0$ then the corresponding $p_i = 0$ and $0 \log 0 = 0$. All the logarithms are considered to the base 2.

In this paper we shall form a functional equation involving a function of two variables which is a generalization of Kendall's equation (1.1) and which gives rise to an inaccuracy function as its solution under suitable boundary conditions. A characterization of Kerridge's inaccuracy function has been given and a study of some important properties of the new measure of inaccuracy has been made.

2. FUNCTIONAL EQUATION IN TWO VARIABLES

Let δ_n denote the set of all complete finite discrete probability distributions, i.e.

$$\delta_n = \left\{ (p_1, \dots, p_n) : p_i \geq 0, \sum_{i=1}^n p_i = 1. \right\}.$$

Given two distributions $P = (p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_n)$, we consider the function $H_n(P; Q)$ taken to satisfy the following postulates:

(I) $H_2(p_1, p_2; q_1, q_2)$ with $(p_1, p_2), (q_1, q_2) \in \delta_2$ and

$H_3(p_1, p_2, p_3; q_1, q_2, q_3)$, with $(p_1, p_2, p_3), (q_1, q_2, q_3) \in \delta_3$ are real valued functions symmetric in the sense that for $n = 2, 3$; $H_n(p_1, \dots, p_n; q_1, \dots, q_n) = H_n(p_{i(1)}, \dots, p_{i(n)}; q_{i(1)}, \dots, q_{i(n)})$ where $(i(1), i(2), \dots, i(n))$ is some permutation of $(1, 2, \dots, n)$.

(II) H_2 and H_3 are connected by the relation

$$H_3 (tp_1, (1-t) p_1, p_2; t' q_1, (1-t') q_1, q_2) = H_2 (p_1, p_2; q_1, q_2) + p_1 H_2 (t, 1-t; t', 1-t')$$

when $t, t' \in (0,1)$ and $(p_1, p_2), (q_1, q_2) \in \delta_2$.

It is easy to see that the function $f(p; q)$ defined by

$$f(p; q) = H_2 (p, 1-p; q, 1-q), p, q \in (0,1) \tag{2.1}$$

satisfies the functional equation

$$f(x_1; y_1) + (1-x_1) f\left(\frac{x_2}{1-x_1}; \frac{y_2}{1-y_1}\right) = f(x_2; y_2) + (1-x_2) f\left(\frac{x_1}{1-x_2}; \frac{y_1}{1-y_2}\right) \tag{2.2}$$

where $x_1, x_2, y_1, y_2 \in [(0, 1), x_1 + x_2 \leq 1, y_1 + y_2 \leq 1]$, in view of above two Postulates.

Definition—Every real valued function $f(x; y)$ of two variables x and y such that $x, y \in [0, 1]$ satisfying the functional equation (2.2) and the boundary conditions.

$$f(0; 0) = f(1; 1); f\left(\frac{1}{2}; \frac{1}{2}\right) = 1, \tag{2.3}$$

is called an inaccuracy function.

The inaccuracy of the discrete probability distribution Q with respect to probability distribution P obtained from an inaccuracy function f is defined by

$$H_n^f (P; Q) = \sum_{i=2}^n s_i f\left(\frac{p_i}{s_i}; \frac{q_i}{t_i}\right) \tag{2.4}$$

where $s_i = p_1 + p_2 + \dots + p_i, t_i = q_1 + q_2 + \dots + q_i$ and $i = 2, 3, \dots$

Remarks: If $x_1 = y_1 = x, x_2 = y_2 = y$ and we take $f(x; x) = f(x)$ then (2.2) is identical with (1.1) and $f(x; x)$ or $f(x)$ is an information function and (2.4) reduces to (1.3). An inaccuracy function and $H_n^f (P; Q)$ are thus the generalizations of the information function and entropy (1.3) respectively.

Kerridge's Inaccuracy Function

It would be noted here that the function

$$f(x; y) = K(x; y) = \begin{cases} -x \log y - (1-x) \log (1-y) & \text{if } x, y \in (0, 1) \\ 0 & \text{if } x = y = 0 \\ & \text{or } x = y = 1 \\ \infty & \text{if } y = 0, x \neq 0, \text{ or } y = 1, x \neq 1, \end{cases} \tag{2.5}$$

satisfies (2.2) and (2.3). This is an inaccuracy function by definition. Since (2.4) reduces to (1.6) in view of (2.5), we shall call $K(x; y)$ as Kerridge's inaccuracy function. This answers the existence of inaccuracy functions.

The existence of inaccuracy function other than that given above can be explored further. In this direction we state the following theorem which may be proved easily.

Theorem 1—Let R be a real valued function defined in $D = (0, 1] \times (0, 1]$ satisfying

$$R(x_1 x_2; y_1 y_2) = x_1 R(x_2; y_2) + x_2 R(x_1; y_1) \tag{2.6}$$

for all (x_1, y_1) and $(x_2, y_2) \in D$ such that

$$R\left(\frac{1}{2}; \frac{1}{2}\right) = \frac{1}{2}. \tag{2.7}$$

Then the function f defined by

$$f(x; y) = R(x; y) + R(1 - x; 1 - y) \tag{2.8}$$

is an inaccuracy function.

3. A CHARACTERIZATION OF KERRIDGE'S INACCURACY FUNCTION

Theorem 2—If a real valued function R satisfying (2.6) and (2.7) is:

- (i) continuous in D and
- (ii) $R\left(1; \frac{1}{2}\right) = 1$,

then

$$R(x; y) = -x \log y$$

and hence $f(x; y)$ defined in (2.8) is Kerridge's inaccuracy function.

PROOF : Consider a function

$$G(x; y) = x R(1/x; 1/y) \text{ for all real } x, y \geq 1. \tag{3.1}$$

Obviously G is continuous in view of (i) above. By taking $x_1 = y_1, x_2 = y_2$ in (2.6) and denoting $R(x; x)$ by $\phi(x)$

we get

$$\phi(x_1 x_2) = x_1 \phi(x_2) + x_2 \phi(x_1). \tag{3.2}$$

Also taking $x = y$ in (3.1) and putting $G(x; x)$ by $g(x)$, we get

$$g(x) = x\phi(1/x), \text{ for all real } x \geq 1. \tag{3.3}$$

These yield

$$g(1/x_1 x_2) = g(1/x_1) + g(1/x_2), \text{ for } 0 < x_1, x_2 \leq 1$$

which is Cauchy's functional equation whose only continuous solution is

$$g(1/x) = A \log x, \text{ for all } 0 < x \leq 1 \tag{3.4}$$

where A is some real constant. Now (3.4) with (2.7) gives

$$G(x; x) = \log x, \text{ for all real } x \geq 1. \tag{3.5}$$

Let m, n, r, s be any positive integers. Setting $x_1 = 1/m, x_2 = 1/n, y_1 = 1/r, y_2 = 1/s$ in (2.6) and utilizing (3.1), we have

$$G(mn; rs) = G(n; s) + G(m; r). \tag{3.6}$$

By suitable substitutions we first obtain that $G(1; x) = \log x$, for all real $x \geq 1$ and then setting $s = n, m = 1$ in (3.6) and using (3.5) we obtain

$$G(n; m) = \log(m), \text{ for all natural numbers } m, n. \tag{3.7}$$

Thus (3.7) and (3.6) for $s = mn$ give

$$m R(1/m; 1/r) = G(m; r) = \log r. \tag{3.8}$$

Now choose any two rational numbers $x, y \in (0, 1)$. Let $x = m/n$ ($m < n$) $y = p/q$ ($p < q$). Setting $x_1 = m/n, x_2 = 1/m, y_1 = p/q$ and $y_2 = 1/p$ in (2.6), using (3.1), (3.8) and continuity of R , we get

$$R(x; y) = -x \log y, \text{ for all } (x, y) \in D. \tag{3.9}$$

Hence

$$f(x; y) = K(x; y), \text{ for all } (x, y) \in D. \tag{3.10}$$

This completes the proof of the theorem.

4. PROPERTIES AND A CHARACTERIZATION OF MEASURE OF INACCURACY

We now come to some properties of the measure of inaccuracy defined in Section 2 and immediately see that if $f(x; y)$ is an inaccuracy function then

$$f(0; 0) = f(1; 1) = 0 \tag{4.1}$$

and

$$f(x; y) = f(1-x; 1-y), \text{ for all } (x, y) \in (0, 1) \times (0, 1) \cup (0, 0) \cup (1, 1). \tag{4.2}$$

Theorem 3—The sequence of functions $H_n^f : S_n \rightarrow \mathcal{R}$ ($n = 2, 3, \dots$) where $S_n = \delta_n \times \delta_n$, \mathcal{R} the set of reals and $H_n^f (P: Q)$, the measure of inaccuracy defined in (2.4); is :

- (i) Normalized : $H_2^f (\frac{1}{2}, \frac{1}{2}; \frac{1}{2}, \frac{1}{2}) = 1$;
- (ii) Expansible : $H_{n+1}^f (p_1, \dots, p_n, 0; q_1, \dots, q_n, 0)$
 $= H_n^f (p_1, \dots, p_n; q_1, \dots, q_n)$;
- (iii) Symmetric : $H_n^f (p_1, \dots, p_n; q_1, \dots, q_n)$
 $= H_n^f (p_{i(1)}, \dots, p_{i(n)}; q_{i(1)}, \dots, q_{i(n)})$

where $(i(1), i(2), \dots, i(n))$ is some permutation of $(1, 2, \dots, n)$;

- (iv) Recursive : $H_{n+1}^f (p_n t, p_n(1-t), p_1, p_2, \dots, p_{n-1}; q_n t', q_n(1-t'), q_1, q_2, \dots, q_{n-1})$
 $= H_n^f (p_1, \dots, p_n; q_1, \dots, q_n) + p_n H_2^f (t, 1-t; t', 1-t')$.

PROOF : The properties (i) and (ii) follow from (2.3) and (2.4) respectively. Now we indicate in short the proofs of the other properties.

(iii) For $n = 2$, we have

$$H_2^f (p_1, p_2; q_1, q_2) = H_n^f (p_2, p_1; q_2, q_1). \tag{4.3}$$

We prove the symmetry for $n > 2$. From (2.4) and (4.3) we have

$$\begin{aligned} &H_n^f (p_1, p_2, \dots, p_n; q_1, q_2, \dots, q_n) \\ &= H_{n-1}^f (p_1 + p_2, p_3, \dots, p_n; q_1 + q_2, q_3, \dots, q_n) \\ &+ (p_1 + p_2) H_2^f \left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}; \frac{q_1}{q_1 + q_2}, \frac{q_2}{q_1 + q_2} \right) \end{aligned} \tag{4.4}$$

where $p_1 + p_2 \neq 0, q_1 + q_2 \neq 0$ and $n = 3, 4, \dots$

By (4.4), H_n^f is invariant under permutations $(p_3, p_4, \dots, p_n; q_3, q_4, \dots, q_n)$ and within $(p_1, p_2; q_1, q_2)$. Now in order to prove its invariance under any permutation within $(p_1, p_2, \dots, p_n; q_1, q_2, \dots, q_n)$ we shall show that

$$\begin{aligned} &H_n^f (p_1, p_2, p_3, p_4, \dots, p_n; q_1, q_2, q_3, q_4, \dots, q_n) \\ &= H_n^f (p_1, p_3, p_2, p_4, \dots, p_n; q_1, q_3, q_2, q_4, \dots, q_n) \end{aligned} \tag{4.5}$$

which in view of (4.4) is equivalent to

$$\begin{aligned} &H_{n-1}^f (p_1 + p_2, p_3, p_4, \dots, p_n; q_1 + q_2, q_3, q_4, \dots, q_n) \\ &+ (p_1 + p_2) H_2^f \left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}; \frac{q_1}{q_1 + q_2}, \frac{q_2}{q_1 + q_2} \right) \\ &= H_{n-1}^f (p_1 + p_3, p_2, p_4, \dots, p_n; q_1 + q_3, q_2, q_4, \dots, q_n) \\ &+ (p_1 + p_3) H_2^f \left(\frac{p_1}{p_1 + p_3}, \frac{p_3}{p_1 + p_3}; \frac{q_1}{q_1 + q_3}, \frac{q_3}{q_1 + q_3} \right) \end{aligned} \tag{4.6}$$

($p_1 + p_3 \neq 0, q_1 + q_3 \neq 0$)

or by (2.4) and (4.3), expression (4.6) is equivalent to

$$\begin{aligned} &f [p_3 / (p_1 + p_2 + p_3); q_3 / (p_1 + q_2 + q_3)] \\ &+ [(p_1 + p_2) / (p_1 + p_2 + p_3)] f [p_2 / (p_1 + p_2); q_2 / (q_1 + q_2)] \\ &= f [p_2 / (p_1 + p_2 + p_3); q_2 / (q_1 + q_2 + q_3)] \\ &+ [(p_1 + p_3) / (p_1 + p_2 + p_3)] f [p_3 / (p_1 + p_3); q_3 / (q_1 + q_3)]. \end{aligned} \tag{4.7}$$

But $p_i, q_i \in (0, 1), i = 1, 2, 3$ and $p_1 + p_2 + p_3, q_1 + q_2 + q_3 \in (0, 1]$ then putting $x_1 = p_3 / (p_1 + p_2 + p_3), x_2 = p_2 / (p_1 + p_2 + p_3), y_1 = q_3 / (q_1 + q_2 + q_3)$ and $y_2 = q_2 / (q_1 + q_2 + q_3)$ in (4.7) we get (2.2) and f being by supposition an inaccuracy function, (2.2) holds for $x_1, x_2, y_1, y_2 \in (0, 1)$ such $x_1 + x_2 \leq 1, y_1 + y_2 \leq 1$. We have thus proved that H_n^f is symmetric in all cases except when some probabilities are zero which can be disregarded as can be seen by (2.4) and (4.1).

(iv) This follows from (4.4) using symmetry of H_n^f for $p_n \neq 0$. If $p_n = 0$ then same is again true by the expansibility of H_n^f taking $q_n = 0$ also.

Lemma—With the notations used in the preceding theorem, we have

$$\begin{aligned} & H_{m-1+n}^f (p_1, p_2, \dots, p_{m-1}, p_m p_{m1}, \dots, p_m p_{mn}; q_1, q_2, \dots, q_{m-1}, q_m q_{m1}, \dots, q_m q_{mn}) \\ &= H_m^f (p_1, p_2, \dots, p_m; q_1, q_2, \dots, q_m) \\ & \quad + p_m H_n^f (p_{m1}, p_{m2}, \dots, p_{mn}; q_{m1}, q_{m2}, \dots, q_{mn}) \end{aligned} \tag{4.8}$$

where $(p_1, \dots, p_m), (q_1, \dots, q_m) \in \delta_m (p_{m1}, \dots, p_{mn}),$
 $(q_{m1}, \dots, q_{mn}) \in \delta_n.$

The proof of the Lemma follows from mathematical induction using the recursive property of H_n^f . (For entropy see Aczel and Daroczy 1975).

Theorem 4—(Strongly additive property) with the notations used in theorem 3,

$$\begin{aligned} & H_{mn}^f \left(\begin{array}{l} p_1 p_{11}, p_1 p_{12}, \dots, p_1 p_{1n}, \dots, p_m p_{m1}, p_m p_{m2}, \dots, p_m p_{mn}; \\ q_1 q_{11}, q_1 q_{12}, \dots, q_1 q_{1n}, \dots, q_m q_{m1}, q_m q_{m2}, \dots, q_m q_{mn} \end{array} \right) \\ &= H_m^f (p_1, p_2, \dots, p_m; q_1, q_2, \dots, q_m) \\ & \quad + \sum_{j=1}^m p_j H_n^f (p_{j1}, p_{j2}, \dots, p_{jn}; q_{j1}, q_{j2}, \dots, q_{jn}) \end{aligned}$$

where $(p_1, p_2, \dots, p_m), (q_1, q_2, \dots, q_m) \in \delta_m, (p_{j1}, p_{j2}, \dots, p_{jn}), (q_{j1}, q_{j2}, \dots, q_{jn}) \in \delta_n$ for every $j = 1, 2, \dots, m.$

The theorem follows by repeated application of (4.8) using the symmetry property.

We now attempt a problem of characterization of $H_n^f(p_1, \dots, p_n; q_1, \dots, q_n)$, that is, we take some properties that a sequence of functions $I_n: S_n \rightarrow \mathcal{A} (n = 2, 3, \dots)$ satisfies and prove that this function $I_n(p_1, \dots, p_n; q_1 \dots q_n)$ is identical with $H_n^f(p_1 \dots p_n; q_1, \dots, q_n)$ is identical with $H_n^f(p_1, \dots, p_n; q_1, \dots, q_n).$

Theorem 5—Let $I_n: S_n \rightarrow \mathcal{A} (n = 2, 3, \dots)$ be a sequence of functions satisfying the following postulates:

(1) $I_3(p_1, p_2, p_3; q_1, q_2, q_3)$ is symmetric function of its arguments (see Section 2, Postulate I);

(2) $I_2(\frac{1}{2}, \frac{1}{2}; \frac{1}{2}, \frac{1}{2}) = 1;$

(3) $I_n(p_1, p_2, \dots, p_n; q_1, q_2, \dots, q_n)$

$= I_{n-1}(p_1 + p_2, p_3, \dots, p_n; q_1 + q_2, q_3, \dots, q_n)$

$+ (p_1 + p_2) I_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}; \frac{q_1}{q_1 + q_2}, \frac{q_2}{q_1 + q_2}\right)$

for $(p_1, \dots, p_n), (q_1, \dots, q_n) \in \delta_n, n = 3, 4, \dots, p_1 + p_2 > 0, q_1 + q_2 > 0.$

Then the function f defined by

$$f(x; y) = I_2(x, 1-x; y, 1-y), \text{ for all } x, y \in [0, 1] \quad (4.9)$$

is an inaccuracy function and further we have

$$I_n(p_1, \dots, p_n; q_1, \dots, q_n) = H_n^f(p_1, \dots, p_n; q_1, \dots, q_n), \quad (4.10)$$

for all $(p_1, \dots, p_n), (q_1, \dots, q_n) \in \delta_n$.

PROOF: (1), (3) and (4.9) together give

$$\begin{aligned} f(1-x_1; 1-y_1) + (1-x_1) f\left(\frac{x_2}{1-x_1}; \frac{y_2}{1-y_1}\right) \\ = f(1-x_2; 1-y_2) + (1-x_2) f\left(\frac{x_1}{1-x_2}; \frac{y_1}{1-y_2}\right) \end{aligned} \quad (4.11)$$

for all $x_1, x_2, y_1, y_2 \in [0, 1], x_1+x_2 \leq 1, y_1+y_2 \leq 1$.

Putting $x_1 = y_1 = 0$ and $x_2 = y_2 = \frac{1}{2}$ in (4.11) we have

$$f(1; 1) = \frac{1}{2} f(0; 0). \quad (4.12)$$

Also from (3), (1) and (4.9), we have

$$f(1; 1) = f(0; 0). \quad (4.13)$$

Next (4.12) and (4.13) yield

$$f(1; 1) = f(0; 0) = 0. \quad (4.14)$$

From (1), (3) (for $n = 3$) and (4.9) we have

$$f(x; y) = f(1-x; 1-y), \text{ for all } x, y \in [0, 1]. \quad (4.15)$$

Now (4.11) with (4.15) gives (2.2) which proves that f is an inaccuracy function. Next (4.10) follows from (3) using (4.9) by induction.

This study deals with the important additive measure. The functional equation (2.2) admits of parametric generalization leading to non-additive measures of inaccuracy which the authors (Sharma and Autar 1972, 1973) have earlier attempted.

Remark : Solutions of functional equation (2.2) under different boundary conditions give rise to Kullback's (1959) relative-information functions (Sharma and Autar 1974, Autar *in press*).

ACKNOWLEDGEMENTS

The authors are thankful to Prof. U. N. Singh, Dean, Faculty of Mathematics, University of Delhi, for the encouragement. One of the authors (R.A.) is indebted to C.S.I.R. for the award of a Senior Research Fellowship.

Thanks are also due to the referee for his favourable and helpful comments.

REFERENCES

- Aczel, J., and Daroczy, Z. (1975). Measures of information and their characterizations. Academic press, N. Y.
- Autar, R. (*in press*). On relative information functions. *Indian J. pure appl. Math.*
- Daroczy, Z. (1970). Generalized information functions, *Information and Control*, **16**, 36-51.
- Kendall, D.G. (1964). Functional equations in information theory. *Z. Wahrs. Verw. Geb.*, **2**, 225-29.
- Kerridge, D.F. (1961). Inaccuracy and inference. *J. Royal Stat. Soc.*, B **23**, 184-94.
- Kullback, S. (1959). Information theory and statistics. John Wiley and Sons, Inc., New York.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, **27**, 379-423.
- Sharma, B.D., and Autar, R. (1972). Generalized functional equation of two variables in information theory. *Revue Auto. Informat. Rech. Operat. (Paris)*, R-3, 85-98.
- (1973). On characterization of a generalized inaccuracy measure in information theory. *J. appl. Prob.*, **10** (No. 2), 464-68.
- (1974). Relative information functions and their type (α, β) generalizations. *Metrika*, **21**, No. 1, 41-50.
- Tverberg, H. (1958). A new derivation of the information function. *Math. Scand.*, **6**, 297-98.