ANNALES
POLONICI MATHEMATICI
XXXV (1977)

On the branching property of entropy

by INDER JEET TANEJA (Delhi, India)

Abstract. A characterization of Shannon's entropy through axioms by forming Jessen–Karpf–Thourup's [9] system of functional equations has been made in first section. While second section deals with a joint characterization of Shannon's entropy and entropy of type β through axioms. Corresponding functional equations studied by Rathie and Kannappan [11] have also been formed.

Introduction. The problem of associating a measure of information with a discrete finite probability distribution $P = (p_1, \ldots, p_n)$, $p_i \geqslant 0$, $\sum_{i=1}^{n} p_i = 1$, started with Shannon's entropy

$$H_n(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i. \tag{1}$$

Characterizations of this measure arising out of natural considerations have been extensively studied by many authors (for details refer Aczél [2]). Faddeev [6] characterized measure (1) by taking the branching property

$$H_n(p_1, p_2, \ldots, p_n) - H_{n-1}(p_1+p_2, p_3, \ldots, p_n) = p_i H_2(p_1/p_i, p_2/p_i), \tag{2}$$

where $p_i = p_1+p_2 > 0$, along with other postulates.

Generalization of measure (1) studied by Havrda and Charvat [8] using in place of the branching property (2) as

$$H_n^{\beta}(p_1, p_2, \ldots, p_n) - H_{n-1}^{\beta}(p_1+p_2, p_3, \ldots, p_n) = p_i^{\beta} H_2^{\beta}(p_1/p_i, p_2/p_i), \quad \beta > 0, \tag{3}$$

where $p_i = p_1+p_2 > 0$. This leads to the entropy of type β given by

$$H_n^{\beta}(p_1, \ldots, p_n) = (2^{1-\beta}-1)^{-1}\Big[\sum_{i=1}^{n} p_i^{\beta} - 1\Big], \quad \beta \neq 1,\ \beta > 0. \tag{4}$$

Quantity (4) has also been studied by Daróczy [4] by taking a functional equation which actually can be shown to follow from the branching property (3) for $n = 3$, together with symmetry. Some functional generalizations of measure (4) have been characterized by Sharma and author [14]. While a joint characterization of measures (1) and (4) through the generalized additivity have been made by author [15].

It is a matter of natural curiosity to examine what are all such measures if we replace $p_i H_2(p_1/p_i, p_2/p_i)$ by a function $\delta_{n-1}(p_1, p_2)$ in (2) and by replacing p_i or p_i^β by a continuous function $f(p_i)$ in (2) or (3). In this paper, we examine these situations.

In first section of the paper, we characterize Shannon's measure (1) by replacing $p_i H_2(p_1/p_i, p_2/p_i)$ by $\delta_{n-1}(p_1, p_2)$, where $\delta_2(p_1, p_2)$ is a continuous function and form Jessen–Karpf–Thourup's [9] system of functional equations by taking certain axioms.

The investigations made in second section show that no new measures arise by replacing p_i or p_i^β by a continuous function $f(p_i)$ in (2) or (3) and the only entropies arising by such a study are those given in (1) or (4).

In what follows we shall take $0 \log 0 = 0$ and all the logarithms are considered to the base 2.

I. Characterization of Shannon's entropy. In this section, we study the basic quantity of information theory known as Shannon's entropy by some axioms. Earlier these axioms have also been considered by Daróczy [3] (refer also Forte and Daróczy [7]). Though the study in this direction has been made extensively but still we are in a position to give yet another characterization for Shannon's entropy. For this, we consider the following axioms:

(I) $H_n(p_1, \ldots, p_n)$ is symmetric function of its arguments;

(II) $H_{n+1}(p_1, \ldots, p_n, 0) = H_n(p_1, \ldots, p_n)$;

(III) $H_n(p_1, \ldots, p_n) - H_{n-1}(p_1+p_2, p_3, \ldots, p_n) = \delta_{n-1}(p_1, p_2)$ $(n \geqslant 3)$, where $\delta_2(p_1, p_2)$ is a continuous function in the region $D = \{(p_1, p_2): p_1, p_2 \geqslant 0, p_1+p_2 \leqslant 1\}$;

(IV) $H_{2n}\big(p_1 q, p_1(1-q), \ldots, p_n q, p_n(1-q)\big) = H_n(p_1, \ldots, p_n) + H_2(q, 1-q)$.

We prove the following theorem based on above axioms:

THEOREM 1. *If* $H_n(p_1, \ldots, p_n), p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1$ *satisfies axioms* (I)–(IV), *then*

$$H_n(p_1, \ldots, p_n) = -\sum_{i=1}^{n} h(p_i), \tag{5}$$

where h satisfies a functional equation

$$h(pq) = ph(q) + qh(p) \tag{6}$$

for all $p, q \in [0, 1]$.

Before proving the theorem, we state two lemmas due to Daróczy [3].

LEMMA 1. *Let* $\Phi(p_1, p_2) = \delta_2(p_1, p_2), (p_1, p_2) \in D$; *then*

$$\delta_{n-1}(p_1, p_2) = \Phi(p_1, p_2). \tag{7}$$

LEMMA 2. *Let* r *be any number lying in* $[0, 1]$; *then*

$$\Phi\big(rq, r(1-q)\big) = r\Phi(q, 1-q) \tag{8}$$

for every $q \in [0, 1]$.

Proof of Theorem 1. Now setting

$$rq = p_1, \quad r(1-q) = p_2,$$

in (8), we get

$$\Phi(p_1, p_2) = (p_1+p_2)\Phi\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right), \tag{9}$$

where $p_1+p_2 > 0$.

Now taking $\dfrac{1}{p_1+p_2} = p_3$ in (9), we get

$$p_3\Phi(p_1, p_2) = \Phi(p_1p_3, p_2p_3). \tag{10}$$

As H_n is symmetric (by axiom (I)), this gives

$$\Phi(p_1, p_2) = \Phi(p_2, p_1). \tag{11}$$

Now consider

$$\begin{aligned}
&\Phi(p_2+p_1, p_3) - \Phi(p_3+p_2, p_1) \\
&= \{H_2(1, 0) + \Phi(p_2, p_1) + \Phi(p_2+p_1, p_3) + \Phi(p_2+p_1+p_3, 1-p_1-p_2-p_3)\} - \\
&\quad -\{H_2(1, 0) + \Phi(p_3, p_2) + \Phi(p_3+p_2, p_1) + \\
&\qquad + \Phi(p_2+p_1+p_3, 1-p_1-p_2-p_3)\} + \Phi(p_3, p_2) - \Phi(p_2, p_1) \\
&= H_4(p_2, p_1, p_3, 1-p_1-p_2-p_3) - H_4(p_3, p_2, p_1, 1-p_1-p_2-p_3) + \\
&\quad + \Phi(p_3, p_2) - \Phi(p_2, p_1) = \Phi(p_3, p_2) - \Phi(p_2, p_1),
\end{aligned}$$

i.e.,

$$\Phi(p_2+p_1, p_3) + \Phi(p_2, p_1) = \Phi(p_3+p_2, p_1) + \Phi(p_3, p_2). \tag{12}$$

Now the continuous solution of the system of functional equations (10)–(12) (refer Jessen–Karpf–Thourup [9]) is given by

$$\Phi(p_1, p_2) = h(p_1+p_2) - h(p_1) - h(p_2), \tag{13}$$

where h satisfies the equation

$$h(pq) = ph(q) + qh(p) \tag{14}$$

for all $p, q \in [0, 1]$.

Now from axiom (III) and Lemma 1, we can write

$$H_n(p_1, \dots, p_n) = \Phi(p_1, p_2) + \Phi(p_1+p_2, p_3) + \dots + \Phi(p_1+p_2+\dots+p_{n-1}, p_n) + H_2(1, 0) \tag{15}$$

(from axiom (IV), we have $H_2(1, 0) = 0$ by taking $q = 1$).

Thus (15) together with (13) gives

$$H_n(p_1, \dots, p_n) = h(1) - \sum_{i=1}^{n} h(p_i),$$

$$= -\sum_{i=1}^{n} h(p_i) \quad (\text{as } h(1) = 0, \text{ from (14)}),$$

which is (5), where h satisfies equation (14).

Characterization of Shannon's entropy. Now, if we suppose that the function h in (14) is continuous, then continuous solution of (14) (refer Aczél [1]) is given by

$$h(p) = Ap\log p, \tag{16}$$

where A is an arbitrary constant.

Thus (5) together with (16) reduces to

$$H_n(p_1, \dots, p_n) = -A\sum_{i=1}^{n} p_i \log p_i. \tag{17}$$

If, we further take the normalizing condition $H_2(\frac{1}{2}, \frac{1}{2}) = 1$, then (17) reduces to Shannon's entropy (1).

Note. The functional equations (10)–(12) have also been used by Aczél [2] for obtaining the Kendall's [10] or Tverberg's [16] functional equation.

II. A joint characterization of Shannon's entropy and entropy of type β. For the purposes of characterizing the measures of information associated with a probability distribution, we take certain axioms. It will be recognized that these are modifications of Faddeev's [6] axioms used for characterizing Shannon's measure.

The information measure $H_n^f(p_1, \dots, p_n)$ of a probability distribution $P = (p_1, \dots, p_n), p_i \geq 0, \sum_{i=1}^{n} p_i = 1$ satisfies the following axioms:

(a) $H_n^f(p_1, \dots, p_n)$ is a continuous function in the region $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$;

(b) $H_n^f(p_1, \dots, p_n)$ is symmetric with respect to its arguments;

(c) $H^f_{n+m-1}(p_1, \ldots, p_{i-1}, v_1, \ldots, v_m, p_{i+1}, \ldots, p_n) = H^f_n(p_1, \ldots, p_n) + f(p_i)H^f_m(v_1/p_i, \ldots, v_m/p_i)$, where $v_k \geqslant 0$, $k = 1, 2, \ldots, m$, $\sum_{k=1}^{m} v_k = p_i > 0$ and f is a continuous function in $[0, 1]$ such that $f(0) = 0$;

(d) $H^f_2(\frac{1}{2}, \frac{1}{2}) = 1$, $H^f_2(1, 0) = 0$.

THEOREM 2. *The entropies determined by axioms* (a)–(d) *are only of one of the following two forms*:

$$H_n(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i, \tag{18}$$

and

$$H_n(p_1, \ldots, p_n) = (2^{1-\beta}-1)^{-1}\Big[\sum_{i=1}^{n} p_i^{\beta}-1\Big], \quad \beta \neq 1,\ \beta > 0. \tag{19}$$

We give some intermediate results based on above axioms in lemmas below:

LEMMA 3. *If* $v_{ij} \geqslant 0$, $j = 1, 2, \ldots, m_i$, $\sum_{j=1}^{m_i} v_{ij} = p_i > 0$, $i = 1, 2, \ldots, n$, $\sum_{i=1}^{n} p_i = 1$, *then*

$$\begin{aligned} &H^f_{m_1+m_2+\ldots+m_n}(v_{11}, \ldots, v_{1m_1}; v_{21}, \ldots, v_{2m_2}; \ldots; v_{n1}, \ldots, v_{nm_n}) \\ &\qquad = H^f_n(p_1, \ldots, p_n) + \sum_{i=1}^{n} f(p_i) H^f_{m_i}(v_{i1}/p_i, \ldots, v_{im_i}/p_i). \end{aligned} \tag{20}$$

This lemma directly follows from axiom (c).

LEMMA 4. *If* $F(n) = H^f_n(1/n, \ldots, 1/n)$, *then*

$$F(n) = A \log n, \quad \textit{when } f(1/n) = 1/n, \tag{21}$$

or

$$F(n) = B[nf(1/n)-1], \tag{22}$$

where f satisfies a functional equation

$$f(1/nm) = f(1/n)f(1/m), \quad f(1/n) \neq 1/n, \tag{23}$$

n, m being arbitrary positive integers and A, B are arbitrary constants.

Proof. Replacing in Lemma 3, m_i by m and $v_{jj} = 1/nm$, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$, where n and m being positive integers, we have

$$F(nm) = F(n) + nf(1/n)F(m). \tag{24}$$

There arise two cases:

Case I. When $f(1/n) = 1/n$. In this case (24) reduces to

$$F(nm) = F(n) + F(m). \tag{25}$$

Also, from the condition $H_2^f(1, 0) = 0$, we have $F(1) = 0$, which gives

(26) $$\lim_{n\to\infty}\left[F(n+1) - \frac{n}{n+1}F(n)\right] = F(1) = 0.$$

The continuous solution of the number theoretic functional equation (25) under condition (26) (refer Daróczy [5], Rényi [12]) is given by (21).

Case II. When $f(1/n) \neq 1/n$. In this case, symmetry in n, m implies

$$F(nm) = F(mn),$$

i.e.,

$$F(n) + nf(1/n)F(m) = F(m) + mf(1/m)F(n),$$

i.e.,

(27) $$\frac{F(n)}{nf(1/n)-1} = \frac{F(m)}{mf(1/m)-1} = B \quad \text{(say)},$$

provided $f(1/n) \neq 1/n$.

Expression (27) gives

$$F(n) = B[nf(1/n)-1], \quad \text{if } f(1/n) \neq 1/n,$$

which is (22).

Now substituting (22) in (24), we obtain (23).

Lemma 5. *The function f in axiom* (c) *satisfies a functional equation given by*

(28) $$f(pq) = f(p)f(q)$$

for all $p, q \in [0, 1]$.

Proof. From axiom (c), we can write

(29) $$H^f_{n+m-1}(p_1, \ldots, p_{i-1}, v_1, \ldots, v_m, p_{i+1}, \ldots, p_n)$$
$$= H^f_{n+1}(p_1, \ldots, p_{i-1}, v_1, \overline{p}, p_{i+1}, \ldots, p_n) + f(\overline{p})H^f_{m-1}(v_2/\overline{p}, \ldots, v_m/\overline{p}),$$
where $\overline{p} = v_2 + \ldots + v_m > 0$
$$= H^f_n(p_1, \ldots, p_n) + f(p_i)H^f_2(v_1/p_i, \overline{p}/p_i) + f(\overline{p})H^f_{m-1}(v_2/\overline{p}, \ldots, v_m/\overline{p}),$$
where $p_i = v_1 + \overline{p} = v_1 + \ldots + v_m$.

Alternatively, we can write again from axiom (c),

(30) $$H^f_{n+m-1}(p_1, \ldots, p_{i-1}, v_1, \ldots, v_m, p_{i+1}, \ldots, p_n)$$
$$= H^f_n(p_1, \ldots, p_n) + f(p_i)H^f_m(v_1/p_i, \ldots, v_m/p_i)$$
$$= H^f_n(p_1, \ldots, p_n) + f(p_i)\{H^f_2(v_1/p_i, \overline{p}/p_i) + f(\overline{p}/p_i)H^f_{m-1}(v_2/\overline{p}, \ldots, v_m/\overline{p})\}$$
$$= H^f_n(p_1, \ldots, p_n) + f(p_i)H^f_2(v_1/p_i, \overline{p}/p_i) + f(p_i)f(\overline{p}/p_i)H^f_{m-1}(v_2/\overline{p}, \ldots, v_m/\overline{p}).$$

Comparing (27) and (30), we get

$$f\left(\frac{\overline{p}}{p_i}\right) = \frac{f(\overline{p})}{f(p_i)}, \quad f(p_i) \neq 0. \tag{31}$$

Finally, (31) together with (23) under the continuity of the function f gives

$$f(pq) = f(p)f(q)$$

for all reals $p, q \in [0, 1]$. This proves the lemma.

Proof of Theorem 2. We prove the theorem for rationals and then continuity axiom (a) gives the result for reals. For this let $p_i = r_i/m$, $i = 1, 2, \ldots, m$, where $\sum_{i=1}^{n} r_i = m$; r_i's and n being positive integers. Then an application of Lemma 3 gives

$$H_m^f(\underbrace{1/m, \ldots, 1/m}_{r_1}; \ldots; \underbrace{1/m, \ldots, 1/m}_{r_n}) = H_n^f(p_1, \ldots, p_n) + \sum_{i=1}^{n} f(p_i) H_{r_i}^f(1/r_i, \ldots, 1/r_i),$$

i.e.,

$$F(m) = H_n^f(p_1, \ldots, p_n) + \sum_{i=1}^{n} f(p_i) F(r_i),$$

i.e.,

$$H_n^f(p_1, \ldots, p_n) = F(m) - \sum_{i=1}^{n} f(p_i) F(r_i). \tag{32}$$

Equation (32) together with (21) gives

$$H_n(p_1, \ldots, p_n) = -A \sum_{i=1}^{n} p_i \log p_i.$$

Again equation (32) together with (22) gives

$$H_n^f(p_1, \ldots, p_n) = B\left[\sum_{i=1}^{n} f(p_i) - 1\right], \tag{33}$$

where f satisfies the functional equation (28).

From axiom (d), we have

$$A = 1 \quad \text{and} \quad B = (2f(\tfrac{1}{2}) - 1)^{-1}.$$

Thus (33) can be written as

$$H_n^f(p_1, \ldots, p_n) = [2f(\tfrac{1}{2}) - 1]^{-1}\left[\sum_{i=1}^{n} f(p_i) - 1\right], \quad f(p) \neq p, \tag{34}$$

where $f(pq) = f(p)f(q)$ for all reals $p, q \in [0, 1]$.

The most general continuous solution of the functional equation (28) (refer Aczél [1]) is given by

(35) $$f(p) = p^{\beta}, \quad \beta > 0,$$

where β ($\neq 1$) is an arbitrary parameter.

Now (34) together with (35) gives (19). While $f(p) = p$ (i.e., $\beta = 1$), we have (18).

A functional equation. Let us take

(36) $$h(p) = H_2^f(p, 1-p), \quad 0 \leqslant p \leqslant 1;$$

then from symmetry, we have

(37) $$h(p) = h(1-p).$$

Again, if we consider the branching property (i.e., axiom (c)) for $n = 3$, this leads to

(38) $$h(p) + f(1-p)h\left(\frac{q}{1-p}\right) = h(q) + f(1-q)h\left(\frac{p}{1-q}\right)$$

for all $p, q \in [0, 1)$ and $p + q \leqslant 1$.

Next, using the branching property for any n, we get

(39) $$H_n^f(p_1, \ldots, p_n) = \sum_{i=2}^{n} f(s_i)h(p_i/s_i),$$

where $s_i = p_1 + \ldots + p_i > 0$; $i = 2, 3, \ldots, n$ and f satisfies a functional equation (refer Lemma 3) given by

(40) $$f(pq) = f(p)f(q)$$

for all reals $p, q \in [0, 1]$.

The functional equation (38) (refer Rathie and Kannappan [11]) has the general continuous solutions given by

(41) $$h(p) = -p\log p - (1-p)\log(1-p), \quad \text{when } f(p) = p,$$

and

(42) $$h(p) = [2f(\tfrac{1}{2}) - 1]^{-1}[f(p) + f(1-p) - 1], \quad \text{if } f(p) \neq p,$$

where f satisfies a functional equation (40).

Now (39) together with (41) gives Shannon's entropy. While (39) together with (42) gives (34), which under the general continuous solution (35) of the functional equation (40) reduces to type β entropy (19).

Author is thankful to Dr. B. D. Sharma, Reader in Mathematics, University of Delhi, Delhi for guidance in the preparation of this paper and discussions at various stages.

References

[1] J. Aczél, *Lectures on functional equations and their applications*, Academic Press, New York 1966.

[2] — *Some applications of functional equation and inequalities to information measures*, C.I.M.E. Report, III Ciclo-La Mendola, 20–28 August 1970, p. 1–20.

[3] Z. Daróczy, *Über eine Charakterisierung der Shannonschen Entropie*, Statistica (Bologna), 27 (1967), p. 199–205.

[4] — *Generalized information functions*, Information and Control 16 (1970), p. 35–51.

[5] — *On the Shannon measure of information*, Selected Transl. in Maths. Stat. Prob. 10 (1972).

[6] D. K. Faddeev, *On the concept of finite probabilistic scheme* (Russian), Uspchi Mat. Nauk. 11 (1956), p. 227–231.

[7] B. Forte and Z. Daróczy, *A characterization of Schannon's entropy*, Boll. U.M.I.(4); N. 4–5 (1968), p. 631–635.

[8] J. Havrda and F. Charvàt, *Quantification method of classification processes, Concept of Structural a-entropy*, Kybernetika 3 (1967), p. 30–35.

[9] B. Jessen, J. Karpf and A. Thorup, *Some functional equations in groups and rings*, Math. Scand. 22 (1968), 257–265.

[10] D. G. Kendall, *Functional equations in information theory*, Z. Wahrs. verw Geb. 2 (1963), p. 225–229.

[11] P. N. Rathie and Pl. Kannappan, *On a functional equation connected with Shannon's entropy*, Funkcialaj Ekvacioj 14 (1971), p. 153–159.

[12] A. Rényi, *On measures of entropy and information*, Proc. 4-th Berk. Symp. Math. Stat. and Prob. 1 (1961), p. 547–561.

[13] C. E. Shannon, *A mathematical theory of communication*, B.S.T.J. 27 (1948), p. 378–423 and 623–656.

[14] B. D. Sharma and I. J. Taneja *Functional measures in information theory*, Funkcialaj Ekvacioj 17 (1974), No. 3, p. 181–191.

[15] I. J. Taneja, *A point characterization of Shannon's entropy and Daróczy's entropy of type β through a functional equation*, Presented at the Conference on "Measures of Information and their Applications" held at Bombay (India), 16–18 August 1974. Also appeared in J. Math. Sci. 10 (1975), p. 69–74.

[16] H. Tverberg, *A new derivation of the information function*, Math. Scand. 6 (1958), p. 297–298.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF DELHI, INDIA

Reçu par le Rédaction la 10. 6. 1975