INFORMATION AND CONTROL 22, 163–171 (1973)

On a Characterization of Directed Divergence*

P. L. KANNAPPAN

Faculty of Mathematics, University of Waterloo, Waterloo, Ontario

AND

P. N. RATHIE

Department of Mathematics, M.R. Engineering College, Jaipur, India

Shannon's entropy was characterized by many authors by assuming different sets of postulates. One other measure associated with Shannon's entropy is directed divergence or information gain. In this paper, a characterization theorem for the measure directed divergence is given by assuming intuitively reasonable postulates and with the help of functional equations.

1. INTRODUCTION

Consider two finite discrete probability distributions

$$P = (p_1, p_2, \ldots, p_n), \qquad Q = (q_1, q_2, \ldots, q_n),$$

with

$$p_i \geqslant 0, \quad \sum_{i=1}^{n} p_i = 1 \qquad \text{with} \qquad q_j \geqslant 0, \quad 1 = \sum_{j=1}^{n} q_j,$$

where the correspondence between the elements of the two distributions are given by their suffices. Then a measure of directed divergence (Kullback, 1959) or information gain (Renyi, 1961) is defined as

$$I_n \begin{pmatrix} p_1, p_2, \ldots, p_n \\ q_1, q_2, \ldots, q_n \end{pmatrix} = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{1.1}$$

Throughout this paper, $\sum$ will stand for the sum $\sum_{i=1}^{n}$; the logarithm will be taken to the base 2. In (1.1), whenever $q_i = 0$, the corresponding $p_i = 0$;

* This work is partially supported by a National Research Council of Canada grant.

and the convention $0.\log 0 = 0 = 0.\log(0/q_i)$ is followed. These conventions are not used in the proof of the characterization theorem in Section 3.

One interpretation for (1.1) may be given as follows. The quantity $-\log p_i$, in communication theory is usually known as the "information content" in the event E_i with probability p_i or the amount of "self information" associated with the event E_i. Thus $\log(p_i/q_i) = \log p_i - \log q_i$ may be taken as the "information gain" in predicting the event E_i. Therefore (1.1) is the average "information gain". The quantity defined in (1.1) is also interpreted as the mean information for discrimation in favor of hypothesis H_1 against hypothesis H_2 in statistical inference problems (Kullback, 1959). Kerridge (1961) interprets (1.1) as a measure of the error made by the observer in estimating a discrete probability distribution as Q which is in fact P. For interpretations of (1.1) in econometric problems see Theil (1967).

Characterizations of (1.1) in arbitrary probability spaces and continuous anologs are given earlier by Campell (1970, 1970A), Kullback and Khairat (1966), Hobson (1969), Rathie and Kannappan (1971) and Reyni (1961).

The object of this paper is to give a characterization theorem for the measure of information defined in (1.1) by assuming reasonable postulates, parallel to that of Shannon's entropy. A method similar to that employed in Kendall (1964) is followed here in arriving at the characterization theorem.

2. Postulates

In this section we give a set of five postulates which will be used in establishing a characterization theorem for (1.1) in the next section. The postulates are:

Postulate 2.1—Recursivity

$$I_n\begin{pmatrix} p_1, \dots, p_n \\ q_1, \dots, q_n \end{pmatrix} = I_{n-1}\begin{pmatrix} p_1 + p_2, p_3, \dots, p_n \\ q_1 + q_2, q_3, \dots, q_n \end{pmatrix} + (p_1 + p_2)\, I_2\begin{pmatrix} p_1/(p_1+p_2), p_2/(p_1+p_2) \\ q_1/(q_1+q_2), q_2/(q_1+q_2) \end{pmatrix},$$

for $p_1 + p_2 > 0$, $q_1 + q_2 > 0$ *and for all* $n = 3, 4, \dots$.

Postulate 2.2—Symmetry

$$I_3\begin{pmatrix} p_1, p_2, p_3 \\ q_1, q_2, q_3 \end{pmatrix} \text{ is symmetric in pairs } \begin{Bmatrix} p_i \\ q_i \end{Bmatrix}\ i = 1, 2, 3.$$

POSTULATE 2.3—Derivative. *Let*

$$f(p, q) = I_2 \begin{pmatrix} p, 1 - p \\ q, 1 - q \end{pmatrix}, \quad \text{for} \quad (p, q) \in J,$$

where $J =]0, 1[\times]0, 1[\cup \{(0, y)\} \cup \{(1, y')\}$ *with* $y \in [0, 1)$ *and* $y' \in]0, 1]$. *Also the function* f *has continuous first partial derivatives with respect to both variables* $p, q \in (0, 1)$.

POSTULATE 2.4—Normalization

$$I_2 \begin{pmatrix} \frac{2}{3}, \frac{1}{3} \\ \frac{1}{3}, \frac{2}{3} \end{pmatrix} = \tfrac{1}{3}.$$

POSTULATE 2.5—Nullity

$$I_2 \begin{pmatrix} p, 1 - p \\ p, 1 - p \end{pmatrix} = 0, \quad \textit{for} \quad p \in (0, 1).$$

Now we discuss the postulates one by one. The Postulate 2.1 suggests the way in which the measures are added up when the union of two mutually exclusive events are considered. In other words, if an event is split into two mutually exclusive events then the measure is a weighted sum with the weights being probabilities as given above. The Postulate 2.2 says that the directed divergence does not depend on the order in which the possible outcomes are labeled. For mathematical purposes the assumption of symmetry for $n = 3$ is enough to establish the theorem. The Postulate 2.3 is a regularity postulate. The Postulate 2.4 is a normalization postulate which fixes the unit for measurement of the directed divergence. The Postulate 2.5 says that if there is no divergence between the two distributions then the measure vanishes.

3. CHARACTERIZATION THEOREM

In this section we prove the following theorem regarding directed divergence (1.1) given by the set of five postulates of the Section 2.

THEOREM. *The function*

$$I_n \begin{pmatrix} p_1, p_2, \ldots, p_n \\ q_1, q_2, \ldots, q_n \end{pmatrix} \quad \textit{for} \quad p_i \geqslant 0, \quad q_i \geqslant 0,$$

$\sum p_i = 1 = \sum q_i$, *satisfying the postulates* 2.1, 2.2, 2.3, 2.4, *and* 2.5 *described in Section* 2, *is the directed divergence given by*

$$I_n\begin{pmatrix} p_1, p_2, ..., p_n \\ q_1, q_2, ..., q_n \end{pmatrix} = \sum p_i \log \frac{p_i}{q_i}. \tag{1.1}$$

Proof. First we will show that I_2 is symmetric. For $n = 3$, Postulate 2.1 gives

$$\begin{aligned} I_3\begin{pmatrix} p_1, p_2, p_3 \\ q_1, q_2, q_3 \end{pmatrix} &= I_2\begin{pmatrix} p_1 + p_2, p_3 \\ q_1 + q_2, q_3 \end{pmatrix} \\ &\quad + (p_1 + p_2)\, I_2\begin{pmatrix} p_1/(p_1 + p_2), p_2/(p_1 + p_2) \\ q_1/(q_1 + q_2), q_2/(q_1 + q_2) \end{pmatrix} \end{aligned} \tag{3.1}$$

for $p_1 + p_2 > 0$, $q_1 + q_2 > 0$. Also from (3.1) for $n = 3$, we have

$$\begin{aligned} I_3\begin{pmatrix} p_2, p_1, p_3 \\ q_2, q_1, q_3 \end{pmatrix} &= I_2\begin{pmatrix} p_2 + p_1, p_3 \\ q_2 + q_1, q_3 \end{pmatrix} \\ &\quad + (p_2 + p_1)\, I_2\begin{pmatrix} p_2/(p_2 + p_1), p_1/(p_2 + p_1) \\ q_2/(q_1 + q_2), q_1/(q_2 + q_1) \end{pmatrix}, \end{aligned} \tag{3.2}$$

for $p_2 + p_1 > 0$, $q_2 + q_1 > 0$,

Hence (3.2), (3.1), and Postulate 2.2 give

$$I_2\begin{pmatrix} p_1/(p_1 + p_2), p_2/(p_1 + p_2) \\ q_1/(q_1 + q_2), q_2/(q_1 + q_2) \end{pmatrix} = I_2\begin{pmatrix} p_2/(p_1 + p_2), p_1/(p_1 + p_2) \\ q_2/(q_1 + q_2), q_1/(q_1 + q_2) \end{pmatrix}, \tag{3.3}$$

for $p_1 + p_2 > 0$, $q_1 + q_2 > 0$, which shows that $I_2\binom{x_1, x_2}{y_1, y_2}$ is symmetric in pairs $\{\binom{x_i}{y_i}\}$, $i = 1, 2$. In particular (3.3) gives

$$f(0, 0) = f(1, 1). \tag{3.4}$$

Next we will obtain an expression for $f(x, y)$. The Postulate 2.2 gives

$$I_3\begin{pmatrix} p_1, p_2, p_3 \\ q_1, q_2, q_3 \end{pmatrix} = I_3\begin{pmatrix} p_2, p_3, p_1 \\ q_2, q_3, q_1 \end{pmatrix} = I_3\begin{pmatrix} p_3, p_1, p_2 \\ q_3, q_1, q_2 \end{pmatrix},$$

which on using Postulate 2.1 for $n = 3$ and the representation in Postulate 2.3 yields the following functional equations,

$$\begin{aligned} f(p_1 + p_2, q_1 + q_2) &+ (p_1 + p_2) f[p_1/(p_1 + p_2), q_1/(q_1 + q_2)] \\ &= f(p_1, q_1) + (1 - p_1) f[p_2/(1 - p_1), q_2/(1 - q_1)] \\ &= f(p_2, q_2) + (1 - p_2) f[p_1/(1 - p_2), q_1/(1 - q_2)], \\ &\text{for } p_1, p_2, q_1, q_2 \in [0, 1),\ p_1 + p_2, q_1 + q_2 \in (0, 1]. \end{aligned} \tag{3.5}$$

Denoting the partial derivative of f with respect to the first variable by f_1 and differentiating partially with respect to p_1 the equation comprised of the first and third lines of (3.5) we have

$$\begin{aligned} &f_1(p_1 + p_2, q_1 + q_2) + f[p_1/(p_1 + p_2), q_1/(q_1 + q_2)] \\ &\qquad + p_2/(p_1 + p_2) f_1[p_1/(p_1 + p_2), q_1/(q_1 + q_2)] \\ &\quad = f_1[p_1/(1 - p_2), q_1/(1 - q_2)], \\ &\qquad \text{for } p_1, q_1 \in (0, 1), p_2, q_2 \in [0, 1) \text{ with } p_1 + p_2, q_1 + q_2 \in (0, 1]. \end{aligned} \tag{3.6}$$

Also differentiating partially with respect to p_2 the equation comprised of the first and second lines of (3.5), we have

$$\begin{aligned} &f_1(p_1 + p_2, q_1 + q_2) + f[p_1/(p_1 + p_2), q_1/(q_1 + q_2)] \\ &\qquad - p_1/(p_1 + p_2) f_1[p_1/(p_1 + p_2), q_1/(q_1 + q_2)] \\ &\quad = f_1[p_2/(1 - p_1), q_2/(1 - q_1)], \\ &\qquad \text{for } p_2, q_2 \in (0, 1), p_1, q_1 \in [0, 1) \text{ with } p_1 + p_2, q_1 + q_2 \in (0, 1]. \end{aligned} \tag{3.7}$$

Hence subtracting (3.7) from (3.6) gives

$$\begin{aligned} &f_1[p_1/(p_1 + p_2), q_1/(q_1 + q_2)] \\ &\quad = f_1[p_1/(1 - p_2), q_1/(1 - q_2)] - f_1[p_2/(1 - p_1), q_2/(1 - q_1)], \\ &\qquad \text{for } p_1, p_2, q_1, q_2 \in (0, 1), p_1 + p_2, q_1 + q_2 \in (0, 1]. \end{aligned} \tag{3.8}$$

Taking $p_1 = xy/(1 + y + xy)$, $p_2 = y/(1 + y + xy)$, $q_1 = uv/(1 + v + uv)$ and $q_2 = v/(1 + v + uv)$ in (3.8) we get

$$\begin{aligned} &f_1[x/(1 + x), u/(1 + u)] \\ &\quad = f_1[xy/(1 + xy), uv/(1 + uv)] - f_1[y/(1 + y), v/(1 + v)] \\ &\qquad \text{for } x, y, u, v \in (0, \infty). \end{aligned} \tag{3.9}$$

Let

$$F(x, u) = f_1[x/(1 + x), u/(1 + u)], \qquad \text{for } x, u \in (0, \infty). \tag{3.10}$$

Then, since f_1 is continuous, so is F.

Then (3.9) with the help of (3.10) gives

$$F(x, u) + F(y, v) = F(xy, uv), \qquad \text{for} \quad x, y, u, v \in (0, \infty). \tag{3.11}$$

Taking $u = v = 1$ in (3.11), we have

$$F(x, 1) + F(y, 1) = F(xy, 1), \qquad \text{for} \quad x, y \in (0, \infty). \tag{3.12}$$

Equation (3.12) is the well known Cauchy equation with the continuous solution given by Aczel (1966),

$$F(x, 1) = a \log x, \; x \in (0, \infty), \tag{3.13}$$

where a is an arbitrary real constant.

Similarly, taking $x = y = 1$ in (3.11), we have

$$F(1, u) = b \log u, \; u \in (0, \infty) \tag{3.14}$$

where b is an arbitrary real constant.

Again, taking $u = 1, y = 1$ in (3.11), we have

$$F(x, v) = F(x, 1) + F(1, v). \tag{3.15}$$

Hence (3.13), (3.14), and (3.15) give

$$F(x, v) = a \log x + b \log v, \qquad \text{for} \quad x, v \in (0, \infty), \tag{3.16}$$

where a and b are arbitrary real constants.

Thus (3.16) and (3.10) give

$$f_1(x, y) = a \log[x/(1-x)] + b \log[y/(1-y)], \qquad \text{for } x, y \in (0, 1), \tag{3.17}$$

which on integration gives

$$f(x, y) = a\{x \log(x/2) + (1-x) \log[(1-x)/2]\} + bx \log[y/(1-y)] + g(y), \tag{3.18}$$

for $x, y \in (0, 1)$, where g is a function of y alone.

Hence (3.18) and (2.5) give

$$\begin{aligned} g(x) = &-a\{x \log(x/2) + (1-x) \log[(1-x)/2)]\} \\ &- bx \log[x/(1-x)], \qquad \text{for} \quad x \in (0, 1). \end{aligned} \tag{3.19}$$

Therefore (3.18) and (3.19) yield

$$\begin{aligned} f(x, y) = &\; a[x \log x + (1-x) \log(1-x) - y \log y - (1-y) \log(1-y)] \\ &+ b(x-y) \log[y/(1-y)], \qquad \text{for} \quad x, y \in (0, 1). \end{aligned} \tag{3.20}$$

Taking $x = \frac{2}{3}, y = \frac{1}{3}$ in (3.20) and utilizing (2.4) we get

$$b = -1. \tag{3.21}$$

Again taking $p_1 = q_1 = \frac{1}{4}$, $p_2 = \frac{2}{3}$, $q_2 = \frac{1}{3}$ in the second and third equation pair of (3.5) for f, that is,

$$\begin{aligned} f(p_1, q_1) + (1 - p_1) f[p_2/(1 - p_1), q_2/(1 - q_1)] \\ = f(p_2, q_2) + (1 - p_2) f[p_1/(1 - p_2), q_1/(1 - q_2)], \end{aligned} \tag{3.22}$$

for $p_1, p_2, q_1, q_2 \in [0, 1)$, $p_1 + p_2, q_1 + q_2 \in (0, 1]$, we have

$$f(\tfrac{1}{4}, \tfrac{1}{4}) + \tfrac{3}{4} f(\tfrac{8}{9}, \tfrac{4}{9}) = f(\tfrac{2}{3}, \tfrac{1}{3}) + \tfrac{1}{3} f(\tfrac{3}{4}, \tfrac{3}{8}). \tag{3.23}$$

Thus (3.23), (3.21), Postulate 2.5, Postulate 2.4 and (3.20) give

$$a = 1. \tag{3.24}$$

Hence (3.20), (3.21) and (3.24) yield

$$f(x, y) = x \log(x/y) + (1 - x) \log[(1 - x)/(1 - y)], \qquad \text{for} \quad x, y \in (0, 1). \tag{3.25}$$

Next we have to determine $f(0, y)$ and $f(1, y)$, where $y \in (0, 1)$.
Taking $p_1 = 0$ in (3.22), we have

$$f(0, q_1) + f[p_2, q_2/(1 - q_1)] = f(p_2, q_2) + (1 - p_2) f[0, q_1/(1 - q_2)], \tag{3.26}$$

for $p_2, q_2 \in (0, 1)$, $q_1 \in [0, 1)$, $q_1 + q_2 \in (0, 1]$.
For $q_1 = \frac{1}{2}$, $q_2 = \frac{1}{4}$, (3.26) reduces to

$$f(0, \tfrac{1}{2}) + f(p_2, \tfrac{1}{2}) = f(p_2, \tfrac{1}{4}) + (1 - p_2) f(0, \tfrac{2}{3}), \qquad \text{for} \quad p_2 \in (0, 1). \tag{3.27}$$

The equation (3.27) for $p_2 = \frac{1}{2}$, with the help of (2.5) gives

$$f(0, \tfrac{1}{2}) = f(\tfrac{1}{2}, \tfrac{1}{4}) + \tfrac{1}{2} f(0, \tfrac{2}{3}). \tag{3.28}$$

Thus (3.28), (3.27) and (3.25) imply

$$f(0, \tfrac{1}{2}) = 1. \tag{3.29}$$

Taking $q_2 = 1 - 2q$, in (3.26) and utilizing (3.29) and (3.25), (3.26) gives

$$f(0, q_1) = -\log(1 - q_1), \qquad \text{for} \quad q_1 \in (0, \tfrac{1}{2}). \tag{3.30}$$

Again putting $q_2 = \frac{1}{2}$ in (3.26) and using (3.25) and (3.30), we get

$$f(0, 2q_1) = -\log(1 - 2q_1), \qquad \text{for} \quad 2q_1 \in (0, 1). \tag{3.31}$$

Hence (3.30) and (3.31) yield

$$f(0, q) = -\log(1 - q), \qquad \text{for} \quad q \in (0, 1). \tag{3.32}$$

The symmetry of I_2 will imply

$$f(p, q) = f(1 - p, 1 - q), \quad \text{for } p, q \in J, \text{ with } f(0, 0) = f(1, 1) = 0, \tag{3.33}$$

[obtained by putting $p_1 = 0 = q_1$, in the last two equations of (3.5)] from which on using (3.32) we get

$$f(1, q) = -\log q, \qquad \text{for} \quad q \in (0, 1). \tag{3.34}$$

Hence (3.25), (3.32), (3.24) and $f(0, 0) = 0 = f(1, 1)$, imply

$$f(x, y) = x \log(x/y) + (1 - x) \log[(1 - x)/(1 - y)], \quad \text{for} \quad (x, y) \in J. \tag{3.35}$$

Now repeated application of Postulate 2.1 gives

$$I_n \begin{pmatrix} p_1, ..., p_n \\ q_1, ..., q_n \end{pmatrix} = \sum_{i=2}^{n} r_i I_2 \begin{pmatrix} r_{i-1}/r_i, p_i/r_i \\ s_{i-1}/s_i, q_i/s_i \end{pmatrix}, \tag{3.36}$$

where $r_i = p_1 + \cdots + p_i$ and $s_i = q_1 + \cdots + q_i$; which on using (3.35) proves the theorem. This completes the proof of this theorem.

RECEIVED: May 15, 1972

REFERENCES

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

CAMPBELL, L. L. (1970), Equivalence of Gauss's principle and minimum discrimination information estimation by probabilities, *Ann. Math. Statist.* **41**, 1011–1015.

CAMPBELL, L. L. (1970A), Characterization of Entropy in Arbitrary Probability Spaces, Queen's University Preprint No. 1970-32, Kingston, Ontario, Canada.

HOBSON, A. (1969), A new theorem of information theory, *J. Statist. Phys.* **1**, 383–391.

KENDALL, D. G. (1964), Functional equations in information theory, *Z. Wahr. Verw. Geb.* **2**, 225–229.

KERRIDGE, D. F. (1961), Inaccuracy and inference, *J. Roy. Statist. Soc.* **23**, 184–194.

KULLBACK, S. (1959), "Information Theory and Statistics," Wiley, New York.

KULLBACK, S., AND LEIBLER, R. A. (1951), On information and sufficiency, *Ann. Math. Statist.* **22**, 79–86.

KULLBACK, S., AND KHAIRAT, M. A. (1966), A note on minimum discrimination information, *Ann. Math. Statist.* **37**, 279–280.

RATHIE, P. N., AND KANNAPPAN, P. L. (1971), On a new characterization of directed-divergence, *in* "Information Theory," 6th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, September 19–25, 1971.

RENYI, A. (1961), On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Statist. Probabil.* 1, 547–561.

THEIL, H. (1967), "Economics and Information Theory," North-Holland, Amsterdam.