ANNALES
POLONICI MATHEMATICI
XXXVIII (1980)

Distributional solutions in information theory, II

by A. KAMIŃSKI (Katowice) and PL. KANNAPPAN (Waterloo, Ontario)

Abstract. The measure generalized directed divergence is characterized with the help of a functional equation, which is solved by means of distributional methods.

1. Let $P = (p_1, \ldots, p_n)$, $Q = (q_1, \ldots, q_n)$, $R = (r_1, \ldots, r_n)$, $p_i, q_i, r_i \geqslant 0$, $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} r_i = 1$, be three discrete probability distributions. Then the measure generalized directed divergence (see [1], [5]–[7]) is given by

$$D_n(P||Q|R) = D_n\begin{pmatrix} p_1, \ldots, p_n \\ q_1, \ldots, q_n \\ r_1, \ldots, r_n \end{pmatrix} = \sum_{i=1}^{n} p_i \log\frac{q_i}{r_i}, \tag{1}$$

where the logarithm base is 2.

In this paper we characterize (1) by a functional equation involving distributions, using a method similar to given in [2]. For similar characterizations of Shannon's entropy, directed divergence and inaccuracy we refer to [3] and [4].

POSTULATES. Let $K_n(P||Q|R)$ be a system of functions defined for $n \geqslant 2$ and satisfying the following axioms:

(i) $K_n(P||Q|R)$ is symmetric in $\begin{pmatrix} p_i \\ q_i \\ r_i \end{pmatrix}$ $(i = 1, \ldots, n)$,

(ii) K_n is continuous,

(iii) $$K_n\begin{pmatrix} p_1, \ldots, p_n \\ q_1, \ldots, q_n \\ r_1, \ldots, r_n \end{pmatrix} = K_{n-1}\begin{pmatrix} p_1+p_2, p_3, \ldots, p_n \\ q_1+q_2, q_3, \ldots, q_n \\ r_1+r_2, r_3, \ldots, r_n \end{pmatrix} + (p_1+p_2)K_2\begin{pmatrix} \frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2} \\ \frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2} \\ \frac{r_1}{r_1+r_2}, \frac{r_2}{r_1+r_2} \end{pmatrix},$$

whenever $p_1+p_2, q_1+q_2, r_1+r_2 > 0$.

THEOREM. *If the functions K_n satisfy conditions* (i)–(iii), *then*

$$K_n(P\,||Q|R) = a\sum_{i=1}^{n} p_i \log p_i + b\sum_{i=1}^{n} p_i \log q_i + c\sum_{i=1}^{n} p_i \log r_i, \tag{2}$$

where a, b, c are arbitrary constants.

Remark. As in [4], it can be shown that finding K_n in (2) is equivalent to solving the functional equation

$$L\begin{pmatrix} x, y, z \\ u, v, w \\ p, q, r \end{pmatrix} = L\begin{pmatrix} x+y, 0, z \\ u+v, 0, w \\ p+q, 0, r \end{pmatrix} + L\begin{pmatrix} x, y, 0 \\ u, v, 0 \\ p, q, 0 \end{pmatrix} \tag{3}$$

with L symmetric and positively homogeneous of order 1 with respect to x, y, z and of order 0 with respect to u, v, w and p, q, r, where

$$L\begin{pmatrix} x, y, z \\ u, v, w \\ p, q, r \end{pmatrix} = (x+y+z)K_3\begin{pmatrix} \frac{x}{x+y+z}, \frac{y}{x+y+z}, \frac{z}{x+y+z} \\ \frac{u}{u+v+w}, \frac{v}{u+v+w}, \frac{w}{u+v+w} \\ \frac{p}{p+q+r}, \frac{q}{p+q+r}, \frac{r}{p+q+r} \end{pmatrix}. \tag{4}$$

Further, instead of solving the functional equation (3) for three triples (x, y, z), (u, v, w), (p, q, r) of variables, we solve a more general one for a system of n triples.

2. Let $\boldsymbol{D}$ be the domain in $\boldsymbol{R}^{3n}$ which consists of all points

$$(x, y, z) = \begin{pmatrix} x_1, y_1, z_1 \\ \cdots\cdots\cdots \\ x_n, y_n, z_n \end{pmatrix} \in \boldsymbol{R}^{3n}$$

such that $x_i, y_i, z_i \geqslant 0$ $(i = 1, \ldots, n)$ and at least two of the columns

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

have all elements positive.

Let $\boldsymbol{D}^0$ be the interior of $\boldsymbol{D}$, i.e., the set of points $(x, y, z) \in \boldsymbol{R}^{3n}$ for which $x_i, y_i, z_i > 0$ $(i = 1, \ldots, n)$.

Let

$$K(x, y, z) = K\begin{pmatrix} x_1, y_1, z_1 \\ \cdots\cdots\cdots \\ x_n, y_n, z_n \end{pmatrix}$$

be a distribution defined on an open set $\mathcal{O}$ containing $\boldsymbol{D}$.

The distribution K is said to be *symmetric* on $\boldsymbol{D}^0$, if

$$K(x, y, z) = K(y, x, z) = K(x, z, y) \tag{5}$$

for $(x, y, z) \in \boldsymbol{D}^0$. Note that the remaining symmetry equalities follow from (5).

The distribution K is said to be $(1, 0, \ldots, 0)$-*homogeneous* on $\boldsymbol{D}^0$, if

$$K\begin{pmatrix}\lambda_1 x_1, \lambda_1 y_1, \lambda_1 z_1\\ \dots\dots\dots\dots\\ \lambda_n x_n, \lambda_n y_n, \lambda_n z_n\end{pmatrix} = \lambda_1 K\begin{pmatrix}x_1, y_1, z_1\\ \dots\dots\\ x_n, y_n, z_n\end{pmatrix}$$

for $\lambda_1, \ldots, \lambda_n > 0$ and $(x, y, z) \in \boldsymbol{D}^0$.

The symbols

$$K(x, y, 0) = K\begin{pmatrix}x_1, y_1, 0\\ \dots\dots\\ x_n, y_n, 0\end{pmatrix} \quad \text{and} \quad K(x+y, 0, z) = K\begin{pmatrix}x_1+y_1, 0, z_1\\ \dots\dots\dots\dots\\ x_n+y_n, 0, z_n\end{pmatrix}$$

will be meant in the sense of generalized operations on distributions (see [8] and [3]).

THEOREM 2. *If a distribution $K(x, y, z)$, defined in an open set O containing $\boldsymbol{D}$, is symmetric and $(1, 0, \ldots, 0)$-homogeneous in $\boldsymbol{D}^0$ and satisfies, in $\boldsymbol{D}^0$, the equation*

$$K(x, y, z) = K(x+y, 0, z) + K(x, y, 0), \tag{6}$$

then it is, in $\boldsymbol{D}^0$, of the form

$$K(x, y, z) = \sum_{i=1}^{n} a_i[(x_1+y_1+z_1)\log(x_i+y_i+z_i) - \\ - x_1 \log x_i - y_1 \log y_i - z_1 \log z_i], \tag{7}$$

where a_i are constants.

Proof. Let

$$F(x_1, \ldots, x_n) = K(1-x, x, 0) = K\begin{pmatrix}1-x_1, x_1, 0\\ \dots\dots\dots\\ 1-x_n, x_n, 0\end{pmatrix}$$

for $x_i \in (0, 1)$ $(i = 1, 2, \ldots, n)$.

By the symmetry of K, we get

$$F(x_1, \ldots, x_n) = F(1-x_1, \ldots, 1-x_n). \tag{8}$$

By the symmetry and $(1, 0, \ldots, 0)$-homogeneity of K, we obtain from (6)

$$(9)\qquad K\begin{pmatrix} x_1, y_1, z_1 \\ \cdots\cdots\cdots \\ x_n, y_n, z_n \end{pmatrix} = (x_1+y_1+z_1)K\begin{pmatrix} \frac{x_1+y_1}{x_1+y_1+z_1}, 0, \frac{z_1}{x_1+y_1+z_1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{x_n+y_n}{x_n+y_n+z_n}, 0, \frac{z_n}{x_n+y_n+z_n} \end{pmatrix} +$$

$$+(x_1+y_1)K\begin{pmatrix} \frac{x_1}{x_1+y_1}, \frac{y_1}{x_1+y_1}, 0 \\ \cdots\cdots\cdots\cdots \\ \frac{x_n}{x_n+y_n}, \frac{y_n}{x_n+y_n}, 0 \end{pmatrix}$$

$$= (x_1+y_1+z_1)F\left(\frac{z_1}{x_1+y_1+z_1}, \ldots, \frac{z_n}{x_n+y_n+z_n}\right) +$$

$$+(x_1+y_1)F\left(\frac{y_1}{x_1+y_1}, \ldots, \frac{y_n}{x_n+y_n}\right),$$

which, by symmetry, results in

$$(10)\qquad F(x_1, \ldots, x_n)+(1-x_1)F\left(\frac{u_1}{1-x_1}, \ldots, \frac{u_n}{1-x_n}\right)$$

$$= F(u_1, \ldots, u_n)+(1-u_1)F\left(\frac{x_1}{1-u_1}, \ldots, \frac{x_n}{1-u_n}\right)$$

for $x_i, u_i \in (0, 1)$ with $x_i+u_i < 1$.

We shall prove, by induction, that if F satisfy equation (10), then for every $k = 1, \ldots, n$

$$(11)\qquad F(s_1, \ldots, s_n) = \sum_{i=1}^{k}(-a_i)[s_1\log s_i+(1-s_1)\log(1-s_i)]+$$

$$+A(s_{k+1}, \ldots, s_n)s_1+B(s_{k+1}, \ldots, s_n),$$

where a_i are number constants and $A(s_{k+1}, \ldots, s_n)$, $B(s_{k+1}, \ldots, s_n)$ are some distributions of $n-k$ variables (which, in the case $k = n$, also become constants).

Before we prove (11), note that (10) and (11) imply the equality (cf. [2])

$$(12)\qquad A(x_{k+1}, \ldots, x_n)-A\left(\frac{x_{k+1}}{1-u_{k+1}}, \ldots, \frac{x_n}{1-u_n}\right)$$

$$= B\left(\frac{u_{k+1}}{1-x_{k+1}}, \ldots, \frac{u_n}{1-x_n}\right).$$

By partially differentiating (10), first with respect to x_1 and the resultant with respect to u_1, applying the substitution

$$s_i = \frac{u_i}{1-x_i}, \quad t_i = \frac{x_i}{1-u_i}, \tag{13}$$

we obtain

$$s_1(1-s_1)\frac{\partial^2}{\partial s_1^2}F(s_1, \ldots, s_n) = \mathrm{const} \qquad (s_i \in (0,1)),$$

so that (cf. [2])

$$F(s_1, \ldots, s_n) = -a_1[s_1 \log s_1 + (1-s_1)\log(1-s_1)] + A(s_2, \ldots, s_n)\cdot s_1 + \\ + B(s_2, \ldots, s_n),$$

where a_1 is a constant and $A(s_2, \ldots, s_n)$, $B(s_2, \ldots, s_n)$ are distributions. This means that (11) is true for $k = 1$.

Now we assume that (11) holds for some k, which implies (12). Differentiating (12) with respect to u_{k+1} and using (13) we obtain

$$-(1-t_{k+1})t_{k+1}\frac{\partial}{\partial t_{k+1}}A(t_{k+1}, \ldots, t_n) = (1-s_{k+1})\frac{\partial}{\partial s_{k+1}}B(s_{k+1}, \ldots, s_n)$$

for $t_{k+1}, \ldots, t_n, s_{k+1}, \ldots, s_n \in (0,1)$. Hence

$$A(t_{k+1}, \ldots, t_n) = -a_{k+1}[\log t_{k+1} - \log(1-t_{k+1})] + A(t_{k+2}, \ldots, t_n) \tag{14}$$

and

$$B(s_{k+1}, \ldots, s_n) = -a_{k+1}\log(1-s_{k+1}) + B(s_{k+2}, \ldots, s_n), \tag{15}$$

where $A(t_{k+2}, \ldots, t_n)$ and $B(s_{k+2}, \ldots, s_n)$ are some distributions and a_{k+1} is a constant.

Now, (14) and (15) together with (11) lead to

$$F(s_1, \ldots, s_n) = \sum_{i=1}^{k+1}(-a_i)[s_1 \log s_i + (1-s_i)\log(1-s_i)] + \\ + A(s_{k+2}, \ldots, s_n)\cdot s_1 + B(s_{k+2}, \ldots, s_n),$$

which, by induction, proves equation (11) for all $k = 1, \ldots, n$.

In particular, for $k = n$ we have

$$F(s_1, \ldots, s_n) = \sum_{i=1}^{n}(-a_i)[s_1 \log s_i + (1-s_1)\log(1-s_i)] + As_1 + B, \tag{16}$$

where A and B are real constants.

By (12), we get $B = A - A = 0$. On the other hand, from (8) and (16) it follows that $A = 0$. Thus

$$F(s_1, \ldots, s_n) = \sum_{i=1}^{n}(-a_i)[s_1 \log s_i + (1-s_1)\log(1-s_i)].$$

Hence, by (9), we obtain (7), which finishes the proof.

3. Proof of Theorem 1. For $n = 3$, we get the solution of the functional equation (3), which together with (4) and (iii) gives the required form K_n of (2) (cf. [3], [4]).

Note that adopting additionally the postulate of normalization (cf. [7]):

$$\text{(iv)}\quad K_2\begin{pmatrix}\frac{1}{2}, \frac{1}{2}\\ \frac{1}{2}, \frac{1}{2}\\ \frac{1}{2}, \frac{1}{2}\end{pmatrix} = 0, \quad K_2\begin{pmatrix}\frac{2}{3}, \frac{1}{3}\\ \frac{2}{3}, \frac{1}{3}\\ \frac{1}{3}, \frac{2}{3}\end{pmatrix} = \frac{1}{3}, \quad K_2\begin{pmatrix}\frac{2}{3}, \frac{1}{3}\\ \frac{1}{3}, \frac{2}{3}\\ \frac{1}{3}, \frac{2}{3}\end{pmatrix} = 0,$$

we obtain, in (2), $a = 0$, $b = 1$ and $c = -1$, i.e.,

$$K_n(P|\,|Q|R) = D_n(P|\,|Q|R)$$

(the generalized directed divergence).

Remark 2. It is easy to deduce from Theorem 2 also the form of Shannon's entropy, directed divergence and inaccuracy (see [4], [3]).

References

[1] J. Aczél and P. Nath, *Axiomatic characterizations of some measures of divergence in information*, Z. Wahr. verw. Geb. 21 (1972), p. 215–224.

[2] A. Kamiński and Pl. Kannappan, *A note on some theorem in information theory*, Bull. Acad. Polon. Sci., Sér. Sci. Math., Astr., Phys. 25 (1977), p. 925–928.

[3] A. Kamiński, Pl. Kannappan and J. Mikusiński, *Distributional solutions in information theory*, I, Ann. Polon. Math. 36 (1979), p. 101–110.

[4] A. Kamiński and J. Mikusiński, *On the entropy equation*, Bull. Acad. Polon. Sci., Sér, Sci. Math., Astr., Phys. 22 (1974), p. 319–323.

[5] Pl. Kannappan, *On a new characterization of generalized directed divergence*, Proc. 7th Prague Conference, 1974.

[6] — *On various characterizations of generalized directed divergence* (to appear).

[7] — and P.N. Rathie, *An axiomatic characterization of generalized directed divergence*, Kybernetika 9 (1973), p. 330–337.

[8] J. Mikusiński, *Iterated operations on distributions*, Bull. Acad. Polon. Sci., Sér. Sci. Math., Astr., Phys. 23 (1975), p. 23–25.

MATHEMATICAL INSTITUTE OF THE POLISH ACADEMY OF SCIENCES
SECTION IN KATOWICE, KATOWICE, POLAND
UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA

Reçu par la Rédaction le 1. 11. 1977