ANNALES
POLONICI MATHEMATICI
XXVI (1972)

An application of a functional equation to information theory

by P. KANNAPPAN and P. N. RATHIE (Waterloo)

1. Introduction. Let $P = (p_1, \ldots, p_n)$, $p_i \geqslant 0$, $\sum_{i=1}^{n} p_i = 1$, $Q = (q_1, \ldots, q_n)$, $q_i \geqslant 0$, $\sum_{i=1}^{n} q_i = 1$ and $R = (r_1, \ldots, r_n)$, $r_i \geqslant 0$, $\sum_{i=1}^{n} r_i = 1$ be three finite discrete probability distributions. Then we define a generalized directed-divergence function of type β ($\beta \neq 1$) and a generalized directed-divergence of type β ($\beta \neq 1$) as follows:

DEFINITION 1. A real valued function f defined on $I \times I \times I$, where $I = [0, 1]$, is called a *generalized directed-divergence function of type* β ($\neq 1$) if f is a solution of the functional equation,

$$(1) \qquad f(x, y, z) + (1-x)(1-y)^{\beta-1}(1-z)^{1-\beta} f\left(\frac{u}{1-x}, \frac{v}{1-y}, \frac{w}{1-z}\right) = f(u, v, w) + (1-u)(1-v)^{\beta-1}(1-w)^{1-\beta} f\left(\frac{x}{1-u}, \frac{y}{1-v}, \frac{z}{1-w}\right)$$

for $x, y, z, u, v, w \in [0, 1)$, $x+u, y+v, z+w \in I$, satisfying further

$$(2) \qquad f(0, 0, 0) = f(1, 1, 1)$$

and

$$(3) \qquad f(1, 1, \tfrac{1}{2}) = f(0, 0, \tfrac{1}{2}) = 1.$$

DEFINITION 2. If f is a generalized directed-divergence function of type β ($\neq 1$) as defined above, then the generalized directed-divergence of type β ($\neq 1$) is defined by the relation

$$(4) \qquad I_n^{\beta}\begin{pmatrix} p_1 & \ldots & p_n \\ q_1 & \ldots & q_n \\ r_1 & \ldots & r_n \end{pmatrix} \stackrel{\text{df}}{=} \sum_{i=2}^{n} P_i Q_i^{\beta-1} R_i^{1-\beta} f\left(\frac{p_i}{P_i}, \frac{q_i}{Q_i}, \frac{r_i}{R_i}\right),$$

where $P_i = p_1 + \ldots + p_i$, $Q_i = q_1 + \ldots + q_i$, $R_i = r_1 + \ldots + r_i$ for $i = 1, 2, \ldots, n$ with $P_n = Q_n = R_n = 1$.

The object of this paper is to solve the functional equation (1) and to find I_n^β and to discuss some connection of these quantities with the existing measures of information. This has been done in subsequent sections to follow. The case $\beta = 1$ will be discussed by us elsewhere.

2. This section deals with the solution of the functional equation (1) under the boundary conditions (2) and (3) and then to utilize this solution in finding the expression of I_n^β.

THEOREM 1. *If f is a solution of the functional equation* (1) *satisfying the additional conditions* (2) *and* (3), *then f has the form*

$$(5)\qquad f(x, y, z) = [xy^{\beta-1}z^{1-\beta} + (1-x)(1-y)^{\beta-1}(1-z)^{1-\beta} - 1](2^{\beta-1}-1)^{-1}$$

for all $x, y, z \in I$ and conversely.

Remark. In (5) we use the convention $0^a = 0$ $(a \neq 0)$ but nowhere in the proof of the theorem $0^a = 0$ is used. Only to put the solution of (1) in the form (5), this notation is used.

Proof. Taking $x = 0$, $y = 0$, $z = 0$ in (1) gives $f(0, 0, 0) = 0$ and hence from (2), we get

$$(6)\qquad f(1, 1, 1) = f(0, 0, 0) = 0.$$

Now replacing u, v, w in (1) by $1-x$, $1-y$ and $1-z$ respectively and applying (6), we have

$$(7)\qquad f(x, y, z) = f(1-x, 1-y, 1-z) \quad \text{for } x, y, z \in (0, 1).$$

With $p = \dfrac{u}{1-x}$, $q = \dfrac{v}{1-y}$, $r = \dfrac{w}{1-z}$, $\xi = 1-x$, $\eta = 1-y$ and $\zeta = 1-z$ in (1), (1) takes the form

$$(8)\qquad f(1-\xi, 1-\eta, 1-\zeta) + \xi\eta^{\beta-1}\zeta^{1-\beta}f(p, q, r)$$
$$= f(p\xi, q\eta, r\zeta) + (1-p\xi)(1-q\eta)^{\beta-1}(1-r\zeta)^{1-\beta} \times f\left(\frac{1-\xi}{1-p\xi}, \frac{1-\eta}{1-q\eta}, \frac{1-\zeta}{1-r\zeta}\right)$$

for all $p, q, r \in I$ and $\xi, \eta, \zeta \in (0, 1]$ such that $p\xi \neq 1$, $q\eta \neq 1$ and $r\zeta \neq 1$.

Putting $\xi = 1$, $\eta = 1$, $\zeta = \frac{1}{2}$, $r = 0$ in (8) and using (3), we get

$$(9)\qquad f(p, q, 0) = [(1-p)(1-q)^{\beta-1} - 1](2^{\beta-1}-1)^{-1} \quad \text{for } p, q \in [0, 1).$$

Taking $\xi = 1$, $\eta = \frac{1}{2}$, $\zeta = 1$ and $q = 0$ in (8) and using (9) for $p = 0$ and $q = \frac{1}{2}$, we have

$$(10)\qquad f(p, 0, r) = [(1-p)(1-r)^{1-\beta} - 1](2^{\beta-1}-1)^{-1} \quad \text{for } p, r \in [0, 1).$$

Again, taking $\xi = \frac{1}{2}$, $\eta = 1$, $\zeta = 1$ and $p = 0$ in (8) and utilizing (9) we get

$$(11)\qquad f(0, q, r) = [(1-q)^{\beta-1}(1-r)^{1-\beta} - 1](2^{\beta-1}-1)^{-1} \quad \text{for } q, r \in [0, 1).$$

With $p = 1$, $q = 1$ and $r = 0$, (8) becomes

$$(12)\quad f(1-\xi, 1-\eta, 1-\zeta)+\xi\eta^{\beta-1}\zeta^{1-\beta}f(1, 1, 0) = f(\xi, \eta, 0)+ +(1-\xi)(1-\eta)^{\beta-1}f(1, 1, 1-\zeta) \quad \text{for } \xi, \eta \in (0, 1), \zeta \in (0, 1].$$

Taking $\zeta = 1$ in (12), in view of (9), we obtain that

$$(13)\quad f(1, 1, 0) = (1-2^{\beta-1})^{-1}.$$

Putting $p = 0$, $q = 0$ and $r = 1$ in (8), we get

$$(14)\quad f(1-\xi, 1-\eta, 1-\zeta)+\xi\eta^{\beta-1}\zeta^{1-\beta}f(0, 0, 1) = f(0, 0, \zeta)+ +(1-\zeta)^{1-\beta}f(1-\xi, 1-\eta, 1) \quad \text{for } \xi, \eta \in (0, 1], \zeta \in (0, 1).$$

Setting $\xi = 1$, $\eta = 1$ in (14) and making use of (10), we have

$$(15)\quad f(0, 0, 1) = (1-2^{\beta-1})^{-1}.$$

For $\zeta = \frac{1}{2}$ in (12) and using (3), (9) and (13) respectively, (12) gives

$$(16)\quad f(1-\xi, 1-\eta, \tfrac{1}{2}) = [2^{\beta-1}\xi\eta^{\beta-1}+2^{\beta-1}(1-\xi)(1-\eta)^{\beta-1}-1](2^{\beta-1}-1)^{-1} \quad \text{for } \xi, \eta \in (0, 1).$$

Now $\zeta = \frac{1}{2}$ in (14), with the help of (3), (15) and (16), gives

$$(17)\quad f(1-\xi, 1-\eta, 1) = [(1-\xi)(1-\eta)^{\beta-1}-1](2^{\beta-1}-1)^{-1} \quad \text{for } \xi, \eta \in (0, 1).$$

Hence, (10), (14), (15) and (17) give us

$$(18)\quad f(1-\xi, 1-\eta, 1-\zeta) = [\xi\eta^{\beta-1}\zeta^{1-\beta}+ +(1-\xi)(1-\eta)^{\beta-1}(1-\zeta)^{1-\beta}-1](2^{\beta-1}-1)^{-1} \quad \text{for } \xi, \eta, \zeta \in (0, 1).$$

With $p = 1$, $q = 0$, $r = 0$, (8) can be written as

$$(19)\quad f(1-\xi, 1-\eta, 1-\zeta)+\xi\eta^{\beta-1}\zeta^{1-\beta}f(1, 0, 0) = f(\xi, 0, 0)+ +(1-\xi)f(1, 1-\eta, 1-\zeta) \quad \text{for } \xi \in (0, 1), \eta, \zeta \in (0, 1].$$

Taking $\eta = 1$, $\zeta = 1$ in (19) and using (9), we get

$$(20)\quad f(1, 0, 0) = (1-2^{\beta-1})^{-1}.$$

With $p = 1$, $q = 0$, $r = 1$, (8) becomes

$$(21)\quad f(1-\xi, 1-\eta, 1-\zeta)+\xi\eta^{\beta-1}\zeta^{1-\beta}f(1, 0, 1) = f(\xi, 0, \zeta)+ +(1-\xi)(1-\zeta)^{1-\beta}f(1, 1-\eta, 1) \quad \text{for } \xi, \zeta \in (0, 1) \text{ and } \eta \in (0, 1].$$

Taking $\eta = 1$ in (21) and using (10) we have

$$(22)\quad f(1, 0, 1) = (1-2^{\beta-1})^{-1}.$$

With $p=0$, $q=1$, $r=1$, (8) becomes

$$f(1-\xi,1-\eta,1-\zeta)+\xi\eta^{\beta-1}\zeta^{1-\beta}f(0,1,1)=f(0,\eta,\zeta)+ \tag{23}$$
$$+(1-\eta)^{\beta-1}(1-\zeta)^{1-\beta}f(1-\xi,1,1) \quad \text{for } \xi\in(0,1],\ \eta,\zeta\in(0,1).$$

For $\xi=1$ in (23) and utilizing (11), we get

$$f(0,1,1)=(1-2^{\beta-1})^{-1}. \tag{24}$$

Taking $p=0$, $q=1$ and $r=1$ in (8) we have

$$f(1-\xi,1-\eta,1-\zeta)+\xi\eta^{\beta-1}\zeta^{1-\beta}f(0,1,0)=f(0,\eta,0)+ \tag{25}$$
$$+(1-\eta)^{\beta-1}f(1-\xi,1,1-\zeta) \quad \text{for } \xi,\zeta\in(0,1],\ \eta\in(0,1).$$

Putting $\xi=1$, $\zeta=1$ in (25) and using (9) with $p=0$, we get

$$f(0,1,0)=(1-2^{\beta-1})^{-1}. \tag{26}$$

Now (23) by the help of (18), (11) and (24) gives

$$f(1-\xi,1,1)=\xi(1-2^{\beta-1})^{-1} \quad \text{for } \xi\in(0,1). \tag{27}$$

On using (26), (9) and (18), (25) gives

$$f(1-\xi,1,1-\zeta)=[(1-\xi)(1-\zeta)^{1-\beta}-1](2^{\beta-1}-1)^{-1} \tag{28}$$
$$\text{for } \xi,\zeta\in(0,1).$$

Equation (19) with the help of (9), (18) and (20) gives

$$f(1,1-\eta,1-\zeta)=[(1-\eta)^{\beta-1}(1-\zeta)^{1-\beta}-1](2^{\beta-1}-1)^{-1} \tag{29}$$
$$\text{for } \eta,\zeta\in(0,1).$$

Now (12), (13), (9) and (18) imply

$$f(1,1,1-\zeta)=[(1-\zeta)^{1-\beta}-1](2^{\beta-1}-1)^{-1} \quad \text{for } \zeta\in(0,1). \tag{30}$$

In (21), on using (18), (10) and (22), we obtain

$$f(1,1-\eta,1)=[(1-\eta)^{\beta-1}-1](2^{\beta-1}-1)^{-1} \quad \text{for } \eta\in(0,1). \tag{31}$$

With $\zeta=1$ in (19) and utilizing (20) and (9), (19) gives

$$f(1,1-\eta,0)=(1-2^{\beta-1})^{-1} \quad \text{for } \eta\in(0,1]. \tag{32}$$

Also (19) with $\eta=1$ on utilizing (20) and (10) gives

$$f(1,0,1-\zeta)=(1-2^{\beta-1})^{-1} \quad \text{for } \zeta\in(0,1]. \tag{33}$$

Taking $\eta=1$ in (14) with (15) and (10) gives

$$f(1-\xi,0,1)=(1-2^{\beta-1})^{-1} \quad \text{for } \xi\in(0,1], \tag{34}$$

while $\xi = 1$ in (14) with (15) and (11) gives

(35) $$f(0, 1-\eta, 1) = (1-2^{\beta-1})^{-1} \quad \text{for } \eta \in (0, 1].$$

Also (25) for $\xi = 1$ on using (26) and (11) gives

(36) $$f(0, 1, 1-\zeta) = (1-2^{\beta-1})^{-1} \quad \text{for } \zeta \in (0, 1],$$

while (25) for $\zeta = 1$ on utilizing (26) and (9) gives

(37) $$f(1-\xi, 1, 0) = (1-2^{\beta-1})^{-1} \quad \text{for } \xi \in (0, 1].$$

Combining (18), (6), (7), (9), (10), (11), (13), (15), (17), (20), (22). (24) and (26) to (37), we conclude that f is given by (5) for all $x, y, z \in I$, It is easy to verify the converse part by straightforward substitution.

THEOREM 2. *The generalized directed-divergence of type* β $(\neq 1)$ *is given by*

(38) $$I_n^\beta \begin{pmatrix} p_1 & \cdots & p_n \\ q_1 & \cdots & q_n \\ r_1 & \cdots & r_n \end{pmatrix} = \Big(\sum_{i=1}^{n} p_i q_i^{\beta-1} r_i^{1-\beta} - 1\Big)(2^{\beta-1}-1)^{-1}.$$

Proof. Substituting the expression for f from (5) in (4) we have

$$I_n^\beta \begin{pmatrix} p_1 & \cdots & p_n \\ q_1 & \cdots & q_n \\ r_1 & \cdots & r_n \end{pmatrix} = (2^{\beta-1}-1)^{-1} \sum_{i=2}^{n} [p_i q_i^{\beta-1} r_i^{1-\beta} + P_{i-1} Q_{i-1}^{\beta-1} R_{i-1}^{1-\beta} - P_i Q_i^{\beta-1} R_i^{1-\beta}]$$

$$= (2^{\beta-1}-1)^{-1} \Big[\sum_{i=2}^{n} p_i q_i^{\beta-1} r_i^{1-\beta} + P_1 Q_1^{\beta-1} R_1^{1-\beta} - P_n Q_n^{\beta-1} R_n^{1-\beta}\Big]$$

$$= (2^{\beta-1}-1)^{-1} \Big[\sum_{i=1}^{n} p_i q_i^{\beta-1} r_i^{1-\beta} - 1\Big],$$

which is (38).

3. Applications to information theory. This section deals with the connection of I_n^β defined in this paper with the existing measures of information.

The expression (38) for $\beta \to 1$ gives

(39) $$I_n^1 \begin{pmatrix} p_1 & \cdots & p_n \\ q_1 & \cdots & q_n \\ r_1 & \cdots & r_n \end{pmatrix} = \sum_{i=1}^{n} p_i \log(q_i/r_i),$$

which is called the *generalized directed-divergence of type unity*. Here the convention $0 \log 0 = 0$ is followed. Also whenever q_i or r_i is zero then the corresponding p_i is also zero and $\log(q_i/r_i)$ is to be taken as $(\log q_i - \log r_i)$ Also (39) for $P \equiv Q$ gives the well-known directed-divergence [2],

$$I_n^1 \begin{pmatrix} p_1 & \cdots & p_n \\ p_1 & \cdots & p_n \\ r_1 & \cdots & r_n \end{pmatrix} = \sum_{i=1}^{n} p_i \log(p_i/r_i). \tag{40}$$

Generalized case of finite discrete probability distributions $P = (p_1, \ldots, p_n)$, $p_i > 0$, $\sum_{i=1}^{n} p_i \leqslant 1$, $Q = (q_1, \ldots, q_n)$, $q_i > 0$, $\sum_{i=1}^{n} q_i \leqslant 1$ and $R = (r_1, \ldots, r_n)$, $r_i > 0$, $\sum_{i=1}^{n} r_i \leqslant 1$, two measures of generalized directed-divergences are defined in [1] as follows:

$$I_1(P||Q|R) = \sum_{i=1}^{n} p_i \log(q_i/r_i) / \sum_{i=1}^{n} p_i, \tag{41}$$

$$I_\alpha(P||Q|R) = (\alpha-1)^{-1} \log_2 \Big[\sum_{i=1}^{n} p_i q_i^{\alpha-1} r_i^{1-\alpha} / \sum_{i=1}^{n} p_i\Big] \quad \text{for } \alpha \neq 1. \tag{42}$$

The motivation for considering (41) and (42) and some axiomatic characterizations are discussed in [1].

Clearly I_n^β given by (38), and I_α given by (42) are connected by the relation

$$I_n^\beta \begin{pmatrix} p_1 & \cdots & p_n \\ q_1 & \cdots & q_n \\ r_1 & \cdots & r_n \end{pmatrix} = (2^{\beta-1}-1)^{-1} [2^{(\beta-1) I_\beta(P||Q|R)} - 1] \tag{43}$$

for $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} r_i = 1$. Therefore the results of this paper also give alternative characterization of (42) for complete probability distributions.

It is interesting to note that for $P \equiv Q$, (38) reduces to

$$I_n^\beta \begin{pmatrix} p_1 & \cdots & p_n \\ p_1 & \cdots & p_n \\ r_1 & \cdots & r_n \end{pmatrix} = (2^{\beta-1}-1)^{-1} \Big(\sum_{i=1}^{n} p_i^\beta r_i^{1-\beta} - 1\Big), \quad \beta \neq 1, \tag{44}$$

which is called a *directed-divergence of type* β and is recently discussed by us in [3]. This quantity (44) has a connection with Rényi's information gain of order β [4]. In our paper [3], a new definition of the directed-divergence function of type β is given and characterized by means of a functional equation.

References

[1] J. Aczél and P. Nath, *Axiomatic characterizations of some measures of divergence in information*, 4th Hawaii International Conference on System Sciences, Univ. of Hawaii, Honolulu, January 1971.

[2] S. Kullback, *Information theory and statistics*, New York 1959.

[3] P. N. Rathie and P. L. Kannappan, *A directed-divergence function of type β* (to appear in Information and Control).

[4] A. Rényi, *On measures of entropy and information*, Proc. Fourth Berkeley Symposium Math. Statist. and Probability, 1 (1961), p. 547-561.

UNIVERSITY OF WATERLOO
WATERLOO

Reçu par la Rédaction le 25. 3. 1971