Indian J. pure appl. Math., **13**(2) : 158–162, February 1982

AN AXIOMATIC CHARACTERISATION OF INACCURACY FUNCTIONS

P. N. ARORA

Department of Mathematics, Dyal Singh College, Lodhi Road, New Delhi 110003

(*Received* 16 *April* 1980; *after revision* 21 *May* 1981)

The object of this paper is to give an axiomatic characterisation of Inaccuracy functions for generalised discrete probability distributions.

1. INTRODUCTION

Let Δ denote the set of all generalised discrete probability distributions. Rényi (1961) defined on a probability space $(\Omega, \mathscr{B}, P)$ i.e. Ω is an abstract set of elementary events. $\mathscr{B}$ a σ-algebra of subsets of Ω containing Ω itself and P a probability measure, that is, a nonnegative countably additive set function defined on $\mathscr{B}$ such that $P(\Omega) = 1$.

For two complete probability distributions

$$P = (p_1, p_2, \ldots, p_N),$$

$$Q = (q_1, q_2, \ldots, q_N),\ p_k > 0, \quad q_k > 0,$$

$$\sum_{k=1}^{N} p_k = \sum_{k=1}^{N} q_k = 1.$$

Kerridge (1961) defined the inaccuracy

$$H(P \parallel Q) = \sum_{k=1}^{N} p_k \log_2 \frac{1}{q_k}$$

which, when $P = Q$, reduces, to Shannon's entropy

$$H(P) = \sum_{k=1}^{N} p_k \log_2 \frac{1}{p_k} \cdot$$

In this paper, we will characterise axiomatically some generalisations $H(P \parallel Q)$.

2. AN AXIOMATIC CHARACTERISATION

Let $H : \Delta \times \Delta \to R$. We assume the following postulates :

Postulate 1 — $p \to H(P \parallel Q)$ remains unchanged if the elements of P and Q are rearranged without disturbing the correspondence between their elements.

Postulate 2 — $p \to H(\{p\} \parallel \{r\}),\ \ 0 < r \leqslant 1$

is a bounded function of p in any arbitrary small subinterval of (0, 1].

Postulate 3 — $r \to H(\{1\} \parallel \{r\})$ is a continuous function of r, $0 < r \leqslant 1$.

Postulate 4 — $H(\{1\} \parallel \{\tfrac{1}{2}\}) = 1$.

Postulate 5 — Let $H(P_1 \parallel Q_1)$, $H(P_2 \parallel Q_2)$, $P_1, P_2 \in \Delta$ $Q_1, Q_2 \in \Delta$ be defined and $P = P_1 * P_2$, $Q = Q_1 * Q_2$, where the symbol $*$ denotes the direct product of the distributions under consideration.

Then

$$H(P \parallel Q) = H(P_1 \parallel Q_1) + H(P_2 \parallel Q_2)$$

where the correspondence between the elements of P and Q is that induced by the correspondence between the elements of P_1, Q_1 and those of P_2, Q_2 respectively.

Postulate 6 — There exists a continuous and strictly increasing function $y = \phi(x)$ defined for all real x such that denoting by $x = \phi^{-1}(y)$, its inverse function, if $H(P_1 \parallel Q_1)$ and $H(P_2 \parallel Q_2)$ are defined with

$$0 < W(P_1) + W(P_2) \leqslant 1, \quad 0 < W(Q_1) + W(Q_2) \leqslant 1$$

and the correspondence between the elements of $P_1 \cup P_2$ and $Q_1 \cup Q_2$ is that induced by the correspondence between the elements of P_1, Q_1 and those of P_2. Q_2 then

$$H(P_1 \cup P_2 \parallel Q_1 \cup Q_2) = \phi^{-1}\left(\frac{W(P_1)\,\phi(H(P_1 \| Q_1)) + W(P_2)\,\phi(H(P_2 \| Q_2))}{W(P_1) + W(P_2)}\right)$$

Postulate 7 — $H(\{p\} \parallel \{r\}) + H(\{q\} \parallel \{r\}) = H(\{p + q\} \parallel \{r\})$

$$p > 0,\ q > 0,\ 0 < r \leqslant 1,\ \ p + q \leqslant 1.$$

Theorem — Let $H(P \parallel Q)$ be defined for all $P \in \Delta$, $Q \in \Delta$ and satisfy Postulates 1, 2, 3, 4, 5, 6 and 7. Then the function ϕ occurring in Postulate 6 is either linear or exponential. In the former case $H(P \parallel Q) = H_1(P \parallel Q)$, where

$$H_1(P \parallel Q) = \frac{\sum_{k=1}^{N} p_k \log_2 (\,\| q_k)}{\sum_{k=1}^{N} p_k}, \ p_k > 0,$$

$$k = 1, 2, \ldots, N, \quad \sum_{k=1}^{N} p_k \leqslant 1, \qquad \ldots(1)$$

whereas in the later case, $H(P \parallel Q) = H_\alpha(P \parallel Q)$, where (Nath 1970)

$$H_\alpha(P \parallel Q) = \frac{1}{1-\alpha} \log_2 \left[\sum_{k=1}^{N} p_k\, q_k^{\alpha-1} \Big/ \sum_{k=1}^{N} p_k \right], \ \alpha \neq 1. \qquad \ldots(2)$$

PROOF : Let

$$f(p, q) = H(\{p\} \| \{q\}), \quad 0 < p \leqslant 1, \quad 0 < q \leqslant 1. \qquad ...(3)$$

By Postulate 7,

$$f(p, r) + f(q, r) = f(p + q, r), p > 0; q > 0, \\ 0 < r \leqslant 1, \quad 0 < p + q \leqslant 1. \qquad ...(4)$$

Define $\phi_r : (0, 1] \rightarrow R$ as

$$\phi_r(p) = f(p, r). \qquad ...(5)$$

Then (4) reduces to

$$\phi_r(p + q) = \phi_r(p) + \phi_r(q), \; 0 < p \leqslant \tfrac{1}{2}, \; 0 < q \leqslant \tfrac{1}{2}, p + q \leqslant 1 \qquad ...(6)$$

which is a Cauchy functional equation (Aczél 1961) on a restricted domain. By Postulate 2, $\phi_r(p)$ is a bounded function of p in any arbitrary small sub-interval of (0, 1]. Hence, (6) admits solutions of the form

$$\phi_r(p) = g(r)p, \; 0 < p \leqslant 1, \; 0 < r \leqslant 1.$$

Consequently,

$$f(p, r) = pg(r), \; 0 < p \leqslant 1, \; 0 < r \leqslant 1. \qquad ...(7)$$

By additivity Postulate 5,

$$f(pq, rs) = f(p, r) + f(q, s), \quad (p, q, r, s \in (0, 1], (0, 1]), \qquad ...(8)$$

from which it follows that

$$f(p, q) = f(p, 1) + f(1, q), \quad (p, q \in (0, 1]). \qquad ...(9)$$

By (7), (9) reduces to

$$pg(q) = pg(1) + g(q), \quad (p, q \in (0, 1]). \qquad ...(10)$$

Putting $q = 1$ it obviously follows that $g(1) = 0$ which implies that $f(p, 1) = 0$. Also (8) gives

$$f(1, rs) = f(1, r) + f(1, s), \quad (r, s \in (0, 1]). \qquad ...(11)$$

By Postulate 3, $r \rightarrow f(1, r)$ is a continuous function of r, $0 < r \leqslant 1$. Hence, (11) admits of continuous solutions of the form

$$f(1, r) = c \log_2 r, \; r \in (0, 1],$$

where c is an arbitrary constant. By Postulate (4), it follows that $c = -1$. Hence $f(1, r) = \log_2(1/r)$. Using (9) and the result $f(p, 1) = 0$, we get

$$f(p, q) = \log_2 (1/q). \qquad ...(12)$$

Let us write

$$P = \{p_1\} \cup \{p_2\} \cup \{p_3\} \cup \dots \cup \{p_N\}$$

$$Q = \{q_1\} \cup \{q_2\} \cup \{q_3\} \cup \dots \cup \{q_N\}.$$

By Postulates 1 and 6, we get

$$H(P \parallel Q) = \phi^{-1}\left(\left(\sum_{k=1}^{N} p_k\, \phi(\log_2(1/q_k))\right)\Big/\left(\sum_{k=1}^{N} p_k\right)\right). \qquad \dots(13)$$

From Postulate 5,

$$H(P * \{a\} \parallel Q * \{b\}) = H(P \parallel Q) + H(\{a\} \parallel \{b\}), \qquad \dots(14)$$
$$0 < a \leqslant 1,\ 0 < b < 1.$$

Using (13), eqn. (14) reduces to

$$\phi^{-1}\left(\left(\sum_{k=1}^{N} p_k\, \phi(\log_2(1/q_k) + \log_2(1/b))\right)\Big/\left(\sum_{k=1}^{N} p_k\right)\right)$$
$$= \phi^{-1}\left(\left(\sum_{k=1}^{N} p_k\, \phi(\log_2(1/q_k))\right)\Big/\left(\sum_{k=1}^{N} p_k\right)\right) + \log_2(1/b),\ 0<b<1. \quad \dots(15)$$

Setting

$$\rho_k = \frac{p_k}{\sum_{k=1}^{N} p_k}, \quad x_k = \log_2 \frac{1}{q_k}, \quad y = \log_2 \frac{1}{b}, \quad \phi_y(x) = \phi(x + y)$$

(15) reduces to

$$\phi^{-1}\left(\sum_{k=1}^{N} \rho_k \phi_y(x_k)\right) = \phi^{-1}\left(\sum_{k=1}^{N} \rho_k \phi(x_k)\right).$$

Since $\phi_y(x)$ and $\phi(x)$ generate the same mean-values, therefore, $\phi_y(x)$ is a linear function of $\phi(x)$, that is,

$$\phi_y(x) = \phi(x + y) = a(y)\, \phi(x) + b(y),\ a(y) \neq 0. \qquad \dots(16)$$

Proceedings as in Rényi (1961), it can be shown that

$$\phi(x) = cx + d \text{ or } \phi(x) = c2^{(1-\alpha)x} + d,\ \alpha \neq 1,\ c \neq 0.$$

d is an arbitrary constant.

The former when put in (13) gives (1), whereas the later when put in (13) gives (2). This complete the proof of the theorem.

3. Comments

(1) Nath (1970), in his Postulate 2, assumes the continuity of

$$p \to H(\{p\} \parallel \{1\}) \text{ and } q \to H(\{1\} \parallel \{q\}), p, q \in (0, 1].$$

In our Postulate 3, we have assumed only the continuity of $q \to H(\{1\} \parallel \{q\})$ and not of $p \to H(\{p\} \parallel \{1\})$. Also, in his Postulate 2, Nath (1970) assumes the continuity of $q \to H(\{1\} \parallel \{q\})$ and $p \to H(\{p\} \parallel \{r\}), p, q \in (0, 1]$. On the other hand, we have assumed only the boundedness of $p \to H(\{p\} \parallel \{r\}$. In an arbitrary small sub-interval of (0, 1]. The other Postulates are common.

Rényi (1961) did not characterise the inaccuracy. Rather, he characterised the entropy and directed divergence (information gain).

Our approach of characterising the inaccuracy is similar to that of Rényi (1961) for characterising information gain.

(2) We have characterised $H(P \parallel Q)$ and $H_\alpha(P \parallel Q)$ only for generalised probability distributions with positive elements. In case P and Q are generalised probability distribution with nonnegative elements such that $0 < \sum_{k=1}^{N} p_k \leqslant 1$, $0 < \sum_{k=1}^{N} q_k \leqslant 1$, then the above proof can be modified accordingly but we have to restrict to $\alpha > 0$. $\alpha \neq 1$.

Acknowledgement

The author is thankful to the refree for his valuable suggestions.

References

Aczél, J. (1966). Lectures on Functional Equations and their Applications. Academic Press, New York.

Kerridge, D. F. (1961). Inaccuracy and Inference. *J. R. Stat. Soc. Series (Methodological)*, **23**.

Nath, P. (1970). An axiomatic characterisation of inaccuracy for discrete generalised probability distribution. *Opsearch*, **7**, 115–33.

Rényi, A. (1961). On measures of entropy and information. *Proc. Fourth Berk. Symposium on Mathematical Statistics and Probability*, **1**, 547–61.