AXIOMATIC CHARACTERIZATIONS OF THE MEASURES OF INACCURACY AND INFORMATION

por

H. Kaufman and P. N. Rathie*

Abstract. In this paper, various measures of inaccuracy and information useful to Statistical Estimation, Inference, Coding Theory, market situations etc. are characterized by a set of suitable postulates in each case. The theory of Functional Equations is used in deriving the characterization theorems.

1. Introduction. We start with a finite generalized probability distribution P, which is a sequence $(p_1, \ldots, p_n)$ with $p_i > 0$, $W(P) = \sum_{i=1}^{n} p_i \leq 1$. $W(P)$ is called the weight of the distribution P. Let $\triangle$ denote the set of all finite discrete generalized probability distributions. This idea of generalized probability distribution was introduced by Rényi [9].

Let $P = (p_1, \ldots, p_n) \in \triangle$ and $Q = (q_1, \ldots, q_n) \in \Delta$ be two generalized probability distributions whose elements are given in one-toone correspondence as determined by their indices. Then the inaccuracy function of order unity and of order α [6] are given respectively by,

$$(1.1) \qquad H_1(P \mid Q) = -\Sigma p_i \log q_i / \Sigma p_i,$$

and

$$(1.2) \qquad H_\alpha(P \mid Q) = (\alpha - 1)^{-1} \log (\Sigma p_i q_i^{1-\alpha} / \Sigma p_i), \; \alpha \neq 1.$$

(1.1) for complete probability distributions was defined earlier in [4]. Throughout this paper, Σ stands for $\sum_{i=1}^{n}$ and logarithms are taken to the base 2.

* Department of Mathematics, McGill University, Montreal, Canada.
This research is supported by National Research Council of Canada Grant No. A 3057.

Also the information functions or directed divergences of order unity and of order α [5,9] are defined respectively as,

$$(1.3) \qquad I(P \mid Q) = \Sigma p_i \log (p_i/q_i) \,/\, \Sigma p_i,$$

and

$$(1.4) \qquad I_\alpha (P \mid Q) = (\alpha - 1)^{-1} \log (\Sigma p_i^\alpha q_i^{1-\alpha} \,/\, \Sigma p_i), \ \alpha \neq 1.$$

Recently, generalizations [6,7] of the inaccuracy functions (1.1) and (1.2) are given in the following forms:

$$(1.5) \qquad H_1^\beta (P \mid Q) = - \Sigma p_i^\beta \log q_i \,/\, \Sigma p_i^\beta,$$

and

$$(1.6) \qquad H_\alpha^\beta (P \mid Q) = (\alpha - 1)^{-1} \log (\Sigma p_i^\beta q_i^{1-\alpha} \,/\, \Sigma p_i^\beta), \ \alpha \neq 1,$$

(1.5) and (1.6) are called the inaccuracies of order unity and type β and of order α and type β respectively.

The generalizations [3] of information functions (1.3) and (1.4) are given earlier as,

$$(1.7) \qquad I_1^\beta (P \mid Q) = \Sigma p_i^\beta \log (p_i \,/\, q_i) \,/\, \Sigma p_i^\beta,$$

and

$$(1.8) \qquad I_\alpha^\beta (P \mid Q) = (\alpha - 1)^{-1} \log (\Sigma p_i^{\alpha+\beta-1} q_i^{1-\alpha} \,/\, \Sigma p_i^\beta), \ \alpha \neq 1;$$

which are called the information of order unity and type β and of order α and type β respectively.

If p_i's are allowed to take zero values and the convention 0log 0 = 0 is followed, then we have to put some conditions on the parameters α and β. Thus, we have to impose the restrictions $\alpha > 0$ in (1.4); $\beta > 0$ in (1.5), (1.6) and (1.7); and $\beta > 0$, $\alpha + \beta - 1 > 0$ in (1.8).

We denote by $R*S$ the direct product of the generalized probability distributions $R = (r_1, \ldots, r_n)$ and $S = (s_1, \ldots, s_m)$; that is the distribution consisting of the sequence $\{r_i s_j\}$ with $i = 1, \ldots, n$; $j = 1, \ldots, m$. Also for $\sum_{i=1}^{n} r_i + \sum_{j=1}^{m} s_j \leq 1$, we define $R \cup S = (r_1, \ldots, r_n, s_1, \ldots, s_m)$. If $\sum_{i=1}^{n} r_i + \sum_{j=1}^{m} s_j > 1$, then $R \cup S$ is not defined. These notations will be used in this paper later on.

Some of the applications of the measures of inaccuracy and information to statistical estimation, inference, coding theory, market situations etc. are indicated in [4], [5], and [9] and in the various references included in these papers.

The object of this paper is to establish four characterization theorems, two each for inaccuracy and information functions including the α-forms. The results are proved by assuming a set of five postulates in each case suiting the particular situation under consideration.

Very recently, a characterization of (1.3) for arbitrary probability spaces by a set of six postulates is given in [2]. Some results for the quantitios given above and more generalized measures of inaccuracy and information may be found in [6, 7, 8].

2. Characterization of inaccuracy functions. In this section a characterization theorem for the inaccuracy functions (1.1) and (1.2) will be proved by using a set of five postulates and another similar characterization theorem for (1.5) and (1.6) is given without proof.

Postulate 1. $H(p \mid 1)$ and $H(1 \mid q)$ are continuous functions of p and q respectively where $p,\ q \in (0,1]$.

Postulate 2. $H(1 \mid \frac{1}{2}) = 1$.

Postulate 3. $H(\frac{1}{2} \mid 1) = 0$.

Postulate 4. If $P = P_1{}^* P_2$ and $Q = Q_1{}^* Q_2$ for $P_1, P_2, Q_1, Q_2 \in \triangle$, then

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

where the correspondence between the elements of P and Q is that induced by the correspondence between the elements of P_1 and Q_1 and those of P_2 and Q_2. Postulate 5. There exists a continuous and strictly monotonic increasing function $y = g(x)$ defined for all real x such that denoting by $x = g^{-1}(y)$ its inverse function; if $P = P_1 \cup P_2$ and $Q = Q_1 \cup Q_2$ where $P, P_1, P_2, Q, Q_1, Q_2 \in \triangle$ and the correspondence between the elements of P and Q is that induced by the correspondence between the elements of P_1 and Q_1 and those of P_2 and Q_2, then we have

$$H(P \mid Q) = g^{-1}\left[\frac{W(P_1)\, g\{H(P_1 \mid Q_1)\} + W(P_2)\, g\{H(P_2 \mid Q_2)\}}{W(P_1) + W(P_2)}\right]$$

In order to prove the characterization theorems for any $n \geq 2$ it is sufficient to take postulates 4 and 5 for

$$P_1 = (p_1), P_2 = (p_2), Q_1 = (q_1), Q_2 = (q_2), \tag{2.1}$$

and the postulate 4 for

$$P_1 = (p_1, p_2), P_2 = (p), Q_1 = (q_1, q_2), Q_2 = (q). \tag{2.2}$$

Then the form of $H(P \mid Q)$ can be easily obtained by induction from $H(p \mid q)$ and $H(p_1, p_2 \mid q_1, q_2)$. The first theorem to be proved is,

THEOREM 1. The functions $H(P \mid Q)$ satisfying the postulates 1, 2, 3, 4 and 5 have exactly two forms given by either (1.1) or (1.2).

Proof. The proof of theorem 1 depends on the following two lemmas.

LEMMA 1. If $H(p \mid q)$ satisfies the postulates 1, 2, 3 and 4 for (2.1), then

$$H(p \mid q) = -\log q. \tag{2.3}$$

Postulate 4 for (2.1) gives

$$H(p_1 p_2 \mid q_1 q_2) = H(p_1 \mid q_1) + H(p_2 \mid q_2). \tag{2.4}$$

Taking $q_1 = q_2 = 1$ in (2.4), we get

$$H(p_1 p_2 \mid 1) = H(p_1 \mid 1) + H(p_2 \mid 1). \tag{2.5}$$

Using the continuity postulate 1, we easily find from [1, p. 41] that the only continuous solution of the functional equation (2.5) is

$$H(p \mid 1) = a \log p, p \in (0,1]. \tag{2.6}$$

Similarly, taking $p_1 = p_2 = 1$ in (2.4), we have

$$H(1 \mid q) = b \log q, \ q \in (0,1]. \tag{2.7}$$

Finally, taking $q_1 = p_2 = 1$, $p_1 = p$ and $q_2 = q$ in (2.4), we get

$$H(p \mid q) = H(p \mid 1) + H(1 \mid q), \tag{2.8}$$

which on using (2.6) and (2.7) yields

$$H(p \mid q) = a \log p + b \log q, \ p, q \in (0,1]. \tag{2.9}$$

Applying the postulates 2 and 3 to (2.9) we find that $a = 0$ and $b = -1$. This proves lemma 1.

Postulate 4 for (2.2) gives,

$$H(p\,p_1, p\,p_2 \mid q\,q_1, q\,q_2) = H(p_1, p_2 \mid q_1, q_2) + H(p \mid q), \tag{2.10}$$

which on using the postulate 5 for (2.1) and the lemma 1 yields

$$g^{-1}\left[\frac{p_1 g(-\log q\,q_1) + p_2 g(-\log q\,q_2)}{p_1 + p_2}\right] = \tag{2.11}$$
$$= g^{-1}\left[\frac{p_1 g(-\log q_1) + p_2 g(-\log q_2)}{p_1 + p_2}\right] + \log(1/q).$$

Now we proceed to prove the following lemma:

LEMMA 2. The continuous and strictly monotonic function $g(x)$ satisfies the functional equation (2.11) if and only if

$$g(x) = Ax + B,\ A \neq 0 \tag{2.12}$$

or

$$g(x) = A\,2^{cx} + B,\ A \neq 0,\ C \neq 0. \tag{2.13}$$

Setting $x = -\log q_1$, $y = -\log q_2$ and $t = -\log q$ in (2.11), we have

$$g^{-1}\left[\frac{p_1 g(x+t) + p_2 g(y+t)}{p_1 + p_2}\right] = \tag{2.14}$$
$$= g^{-1}\left[\frac{p_1 g(x) + p_2 g(y)}{p_1 + p_2}\right] + t.$$

The only continuous and monotonic solutions [1, p. 153] of the functional equation (2.14) are linear and exponential functions given by (2.12) and (2.13). Thus lemma 2 is proved. Hence from postulate 5 for (2.1) with $g(x = Ax + B,\ A \neq 0$, we have

$$H(p_1, p_2 \mid q_1, q_2) \equiv H_1(p_1, p_2 \mid q_1, q_2) = -(p_1 \log q_1 + p_2 \log q_2) / (p_1 + p_2) \tag{2.15}$$

and with $g(x) = A\,2^{(\alpha-1)x} + B,\ A \neq 0,\ \alpha \neq 1$, we have

$$H(p_1, p_2 \mid q_1, q_2) \equiv H_\alpha(p_1, p_2 \mid q_1, q_2) = \tag{2.16}$$
$$= (\alpha - 1)^{-1} \log\left[(p_1 q_1^{1-\alpha} + p_2 q_2^{1-\alpha}) / (p_1 + p_2)\right],\ \alpha \neq 1.$$

It can be easily seen that (1.1) and (1.2) are obtained from postulate 5, (2.4), (2.15) and (2.16) by induction for any $n \geq 2$. This completes the proof of theorem 1.

Now we take the following postulate instead of postulate 5.

Postulate 6. If $W_\beta(P) = \sum_{i=1}^{n} p_i^\beta$, for $P = (p_1, \ldots, p_n)$ etc., then

$$H(P \mid Q) = g^{-1}\left[\frac{W_\beta(P_1)\, g\{H(P_1 \mid Q_1)\} + W_\beta(P_2)\, g\{H(P_2 \mid Q_2)\}}{W_\beta(P_1) + W_\beta(P_2)}\right].$$

Then the following theorem can be similarly proved for the inaccuracies (1.5) and (1.6).

THEOREM 2. The functions $H(P \mid Q)$ satisfying the postulates 1, 2, 3, 4 and 6 have exactly two forms, $H_1^\beta(P \mid Q)$ and $H_\alpha^\beta(P \mid Q)$ given by (1.5) and (1.6) respectively.

3. Characterization of information functions. This section deals with a characterization of information functions (1.3) and (1.4) by the help of five postulates. Towards the end of this section another similar theorem which characterizes (1.7) and (1.8) is given without proof.

Postulate 7. $I(p \mid 1)$ and $I(1 \mid q)$ are continuous functions of p and q respectively for $p, q \in (0,1]$.

Postulate 8. $I(1 \mid \tfrac{1}{2}) = 1$.

Postulate 9. $I(\tfrac{1}{2} \mid \tfrac{1}{2}) = 0$.

Postulate 10. $I(P \mid Q) = I(P_1 \mid Q_1) + I(P_2 \mid Q_2)$.

Postulate 11.

$$I(P \mid Q) = g^{-1}\left[\frac{W(P_1)\, g\{I(P_1 \mid Q_1)\} + W(P_2)\, g\{I(P_2 \mid Q_2)\}}{W(P_1) + W(P_2)}\right].$$

The same convention regarding the various probability distributions as given in the last section are followed here. The theorem to be proved is:

THEOREM 3. The functions $I(P \mid Q)$ satisfying the postulates 7, 8, 9, 10 and 11 are given by either (1.3) or (1.4).

Proof. The proof of theorem 3 is very much similar to that given in the last section. As before, it is easy to show that,

$$I(p \mid q) = a \log p + b \log q, \; p, q \in (0,1]. \tag{3.1}$$

Now postulates 8 and 9 determine the constants in (3.1) giving $a = 1$ and $b = -1$. Thus

$$I(p \mid q) = \log (p/q). \tag{3.2}$$

The postulates 10 for (2.2) and postulate 11 for (2.1) together with (3.2) and following the procedure of the last section give

$$g^{-1}\left[\frac{p_1 g(x+t) + p_2 g(y+t)}{p_1 + p_2}\right] = \tag{3.3}$$

$$= g^{-1}\left[\frac{p_1 g(x) + p_2 g(y)}{p_1 + p_2}\right] + t,$$

where $x = \log (p_1/q_1)$, $y = \log (p_2/q_2)$ and $t = \log (p/q)$.

As before, the solutions of (3.3) are of the forms given in (2.12) and (2.13). Therefore from postulate 11 for (2.1) when $g(x) = Ax + B$, $A \neq 0$, we have

$$I(p_1, p_2 \mid q_1, q_2) \equiv I_1(p_1, p_2 \mid q_1, q_2) \tag{3.4}$$

$$= -\{p_1 \log (p_1/q_1) + p_2 \log (p_2/q_2)\} / (p_1 + p_2)$$

and when $g(x) = A\, 2^{(\alpha-1)x} + B$, $A \neq 0$, $\alpha \neq 1$, we have

$$I(p_1, p_2 \mid q_1, q_2) \equiv I_\alpha(p_1, p_2 \mid q_1, q_2) \tag{3.5}$$

$$= (\alpha - 1)^{-1} \log \{(p_1^\alpha q_1^{1-\alpha} + p_2^\alpha q_2^{1-\alpha}) / (p_1 + p_2)\}, \; \alpha \neq 1.$$

Thus theorem 3 is proved for $n = 2$. The method of induction proves theorem 3 for any n on using the postulate 11, (3.2), (3.4) and (3.5).

Let us take the following postulate instead of the postulate 11.

Postulate 12.

$$I(P \mid Q) = g^{-1}\left[\frac{W_\beta(P_1)\, g\{I(P_1 \mid Q_1)\} + W_\beta(P_2)\, g\{I(P_2 \mid Q_2)\}}{W_\beta(P_1) + W_\beta(P_2)}\right].$$

Then theorem 4 given below can be proved on lines similar to that of theorem 3.

Theorem 4. The functions $I(P \mid Q)$ satisfying the postulates 7, 8, 9, 10 and 12 are $I_1^\beta(P \mid Q)$ and $I_\alpha^\beta(P \mid Q)$ given by (1.7) and (1.8) respectively.

REFERENCES

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

2. L. L. CAMPBELL, *Characterization of entropy in arbitrary probability spaces*, Queenis University Preprint No. 1970-32, Kingston, Ontario, Canada.

3. J. N. KAPUR, *Information of order α and type β*, Proc. Indian Acad. Sci., A68 (1968), pp. 65-75.

4. D. F. KERRIDGE, *Inaccuracy and inference*, J. Roy. Statist. Soc., B 23 (1961), pp. 184-194.

5. S. KULLBACK, *Information theory and statistics*, John Wiley and Sons, 1959.

6. P. N. RATHIE, *On generalized entropies, inaccuracies and information*, Cahiers Centre Études Recherche Opér. 13 (1971), 98-105.

7. P. N. RATHIE, *On generalized measures of inaccuracy*, information and errors in information, Statistica, 30 (1970), pp. 340-349.

8. P. N. RATHIE, *Some characterization theorems for generalized measures of uncertainty and informarion*, Metrika, 20 (1973), 122-130.

9. A. RÉNYI, *On measures of entropy and information*, Proc. Fourth Berkeley Symposium Math. Statis. and Probability, 1 (1971), pp. 547-561.