UNIQUENESS THEOREMS OF GENERALIZED
ENTROPY AND INFORMATION

BY BHU DEV SHARMA

(Department of Mathematics, Institute of Post-Graduate Studies, University of Delhi, Delhi-7, India)

Received February 10, 1969

(Communicated by Prof. R. S. Varma, F.A.Sc.)

1. INTRODUCTION

DAROCZY⁴ and Aczel⁸ have given joint characterization of entropies of Shannon and Renyi for generalized distributions. The characterization of amount of information has been studied by Renyi.¹⁰ The problem of characterization is half done if we first characterize the self information of an event happening with probability p, instead of coming to entropy of the whole distribution.

In this paper we shall prove the uniqueness theorems of generalized entropies and generalized amount of informations.

2. UNIQUENESS OF GENERALIZED ENTROPIES

We start with self information and postulates on it. Let H (p) be the self information of an event whose probability of happening is p, then H (p) should satisfy the following set of postulates:

A.1 H (p) ≥ 0 is a continuous function of p, 0 < p ≤ 1.
A.2 H (pq) = H (p) + H (q).
A.3 H (½) = 1.

We have the following well-known result:

Lemma 1. Under postulates A.1 to A.3,

\[ H (p) = - \log_a p. \] (1)

Proof.—From A.1 and A.2, it can be easily seen that

\[ H (p) = - c \log p. \]

where the base of the logarithm is arbitrary and c is some positive, arbitrary constant independent of p.

Now using A.3 we immediately get (1).
The entropy of a function is taken as some weighted mean of the self informations of various events with weights as functions of the probabilities of events. However this weighted mean should satisfy certain postulates. Therefore the problem can now be formulated as follows:

Given a probability distribution \( P = (p_1, \ldots, p_n) \) with
\[
\sum_{i=1}^{n} p_i \leq 1,
\]
the entropy of \( P \) is (refer Daroczy\(^5\))
\[
H(P) = \psi^{-1}\left( \frac{\sum_{i=1}^{n} f(p_i) \psi(-\log_2 p_i)}{\sum_{i=1}^{n} f(p_i)} \right), \sum_{i=1}^{n} f(p_i) \leq 1 \tag{2}
\]
where

\textit{B.1} \( \psi \) is a strictly monotonic and continuous function,
\textit{B.2} \( f \) is a positive valued and bounded function defined in \([0, 1]\) such that
\[
f(xy) = f(x)f(y)
\]
\textit{B.3} \( H(P^*Q) = H(P) + H(Q) \)

where for
\[
P = (p_1, \ldots, p_n), \quad Q = (q_1, \ldots, q_m)
\]
\[
P^*Q = (p_1q_1, \ldots, p_q q_m, \ldots, p_n q_1, \ldots, p_n q_m).
\]
\textit{B.4}
\[
H(P \cup Q) = \psi^{-1}\left( \frac{\sum_{i=1}^{n} f(p_i) \psi(H(P)) + \sum_{j=1}^{m} f(q_j) \psi(H(Q))}{\sum_{i=1}^{n} f(p_i) + \sum_{j=1}^{m} f(q_j)} \right)
\]
provided that
\[
\sum_{i=1}^{n} f(p_i) + \sum_{j=1}^{m} f(q_j) \leq 1
\]

We now prove the following uniqueness theorem of generalized entropies,
Theorem 1.—The function $H(P)$ given in (2) under the postulates B.1 to B.4 can have only one of the following two forms:

\[ H(P) = H_1f(P) = -\frac{\sum_{i=1}^{n} f(p_i) \log_2 p_i}{\sum_{i=1}^{n} f(p_i)} \] \hspace{1cm} (3)

or

\[ H(P) = H_\alpha f(P) = \frac{1}{1-\alpha} \log_2 \frac{\sum_{i=1}^{n} f(p_i) p_i^{\alpha-1}}{\sum_{i=1}^{n} f(p_i)} , \alpha \neq 1. \] \hspace{1cm} (4)

Remark.—From postulate B.2, a continuous $f(x)$ can be only of the form $x^\beta$ where $\beta$ is some constant. The quantities (3) and (4) where $f(x) = x^\beta$ occur in the works of Aczel and Daroczy and Kapur, while the case $\beta = 1$ is studied by Renyi.

Proof of Theorem 1.—We have from (1) or by taking $n = 1$ in (2),

\[ H(p) = -\log_2 p. \]

Using B.3, we can write

\[ H(p_1q, p_2q, \ldots, p_nq) = H(p_1, p_2, \ldots, p_n) + H(q) \]

i.e.,

\[ \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(p_i q) \psi(-\log_2 p_i q)}{\sum_{i=1}^{n} f(p_i q)} \right) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(p_i) \psi(-\log_2 p_i)}{\sum_{i=1}^{n} f(p_i)} \right) - \log_2 q \]

i.e.,

\[ \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(p_i) \psi(-\log_2 p_i - \log_2 q)}{\sum_{i=1}^{n} f(p_i)} \right) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(p_i) \psi(-\log_2 p_i)}{\sum_{i=1}^{n} f(p_i)} \right) - \log_2 q. \]

(5)

Set now $-\log_2 p_i = x_i$ and $-\log_2 q = y$, then (5) becomes

\[ \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(2^{-x_i}) \psi(x_i + y)}{\sum_{i=1}^{n} f(2^{-x_i})} \right) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(2^{-x_i}) \psi(x_i)}{\sum_{i=1}^{n} f(2^{-x_i})} \right) + y \]
where

$$\sum_{i=1}^{n} f(2^{-x_i}) \leq 1.$$  

Putting now $\psi(x_i + y) = \psi_y(x_i)$, the above equation reduces to

$$\psi_y^{-1} \left( \frac{\sum_{i=1}^{n} f(2^{-x_i}) \psi_y(x_i)}{\sum_{i=1}^{n} f(2^{-x_i})} \right) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(2^{-x_i}) \psi(x_i)}{\sum_{i=1}^{n} f(2^{-x_i})} \right). \quad (6)$$

This functional equation (refer7 Theorem 83, p. 66 or1 Lemma 1, p. 155), for $\psi_y(x)$ and $\psi(x)$ strictly monotonic and continuous, has following relationship between $\psi_y(x)$ and $\psi(x)$:

$$\psi_y(x) = A\psi(x) + B$$

where $A \neq 0$ and $B$ are arbitrary constants as regards $x$.

In general here

$$\psi_y(x) = \psi(x + y) = A(y)\psi(x) + B(y). \quad (7)$$

Aczel,1 Section 3.13, p. 148, for $\psi(x)$ strictly monotonic function gives that there are exactly two solutions of (7) given by

$$\psi(x) = ax + b, \quad a \neq 0 \quad (8)$$

and

$$\psi(x) = ae^{cx} + b, \quad a \neq 0, \quad c \neq 0. \quad (9)$$

The two forms (3) and (4) correspond to the two solutions above when $c = (1 - a) \log_{e}2$.

This completes the proof of the theorem.

3. **Uniqueness of Generalized Amount of Information**

In this section we shall take up the characterization of the generalized amount of information or of generalized Kullback Leibler9 information. Here also we shall first characterize *self-relative information* or the information conveyed by the happening of an event with probability $q$ while its original probability is $p$. Let us denote this by $I(q/p)$ and this be only a function of $p$ and $q$. We now want $I(q/p)$ to satisfy following postulates:
Uniqueness Theorems of Generalized Entropy and Information

C.1. \( I(q/p) \geq 0 \) if \( q \geq p \) and \( I(q/p) \leq 0 \) if \( q \leq p \) where \( 0 < p, q \leq 1 \).

C.2. \( I(qq'/pp') = I(q/p) + I(q'/p') \)

C.3. \( I(1/2) = 1 \).

Note.—Compare from Fano\(^6 \) §2–5 p. 31.

We now have the following lemma:

**Lemma 2.** Under postulates C.1, C.2 and C.3,

\[
I(q/p) = \log_2 \frac{q}{p}. \tag{10}
\]

**Proof.**—The proof is on the lines of Renyi.\(^{10} \)

Let

\[
f(q, p) = I(q/p),
\]

then from C.2,

\[
f(qq', pp') = f(q, p) + f(q', p'). \tag{11}
\]

Putting

\[
q = q' = 1, f(1, pp') = f(1, p) + f(1, p'). \tag{12}
\]

Now setting \( q = p' = 1 \), and \( q' = q \), we get

\[
f(q, p) = f(1, p) + f(q, 1). \tag{13}
\]

From C.1, \( I(p/p) = 0 \), *i.e.*, \( f(1, p) + f(p, 1) = 0 \).

\[
\therefore (13) \text{ gives } f(q, p) = f(1, p) - f(1, q). \tag{14}
\]

From C.1, \( f(1, p) \) is a decreasing and non-negative function of \( p \) and from (12) it is additive, therefore

\[
f(1, p) = c \log \frac{1}{p}
\]

where \( c \) is a positive constant not depending on \( p \), and the base of the logarithm is arbitrary.
However using C.3,
\[ f(1, p) = \log_2 \frac{1}{p} \]
and therefore
\[ f(q, p) = f(1, p) - f(1, q) = \log_2 \frac{q}{p}. \]

After proving this lemma, we come to the amount of information which we shall take as some weighted mean of the self-relative informations, \( I(p/q) \), where weights are some functions of the probabilities \( q \). Thus we can formulate the problem as follows:

Given a distribution \( Q = (q_1, \ldots, q_n) \) of a set of events whose distribution after some experiment is \( P = (p_1, \ldots, p_n) \) such that
\[ \sum_{i=1}^{n} q_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} p_i \leq 1, \]
the amount of information of \( Q \) relative to \( P \) is
\[ I(Q/P) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(q_i) \psi \left( \log_2 \frac{q_i}{p_i} \right)}{\sum_{i=1}^{n} f(q_i)} \right), \quad \sum_{i=1}^{n} f(q_i) \leq 1 \] (15)

where

D.1. \( \psi \) is a strictly monotonic and continuous function,

D.2. \( f \) is a positive valued, bounded, weight function defined in \([0, 1]\), such that \( f(xy) = f(x) \cdot f(y) \)

D.3. \( I(Q' Q'/P' P') = I(Q/P) + I(Q'/P') \)

the correspondence in \( Q' Q' \) and \( P' P' \) being that induced by the indices of its elements.

D.4. If
\[ \sum_{i=1}^{n} f(q_i) + \sum_{j=1}^{n} f(q_j') \leq 1, \]
\[ I(Q \cup Q'/P \cup P') = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(q_i) \psi(I(Q|P)) + \sum_{j=1}^{m} f(q_j') \psi(I(Q'|P'))}{\sum_{i=1}^{n} f(q_i) + \sum_{j=1}^{m} f(q_j')} \right). \]
We now give the following uniqueness theorem for generalized amount of Informations:

**Theorem 2.**—The function \( I(Q/P) \) under postulates D.1 to D.4 can be only of one of the following two forms:

\[
I(Q|P) = I_f(Q|P) = \frac{\sum_{i=1}^{n} f(q_i) \log_2 \frac{q_i}{p_i}}{\sum_{i=1}^{n} f(q_i)} \tag{16}
\]

or

\[
I(Q|P) = I_a f(Q|P) = \frac{1}{a - 1} \log_2 \frac{\sum_{i=1}^{n} f(q_i) \left(\frac{q_i}{p_i}\right)^{a-1}}{\sum_{i=1}^{n} f(q_i)} , a \neq 1. \tag{17}
\]

**Remark.**—Quantities (16) and (17) when \( f(x) = x \) have been studied by Renyi and when \( f(x) = x^\beta \) have been studied by Sharma.11 For study of (15) see Sharma.12

**Proof of Theorem 2.**—We shall only briefly indicate the proof which runs on the lines of proof of theorem 1.

We have here corresponding to (5),

\[
\psi^{-1} \left( \frac{\sum_{i=1}^{n} f(q_i) \psi \left( \log_2 \frac{q_i}{p_i} + \log_2 \frac{q}{p} \right)}{\sum_{i=1}^{n} f(q_i)} \right) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(q_i) \psi \left( \log_2 \frac{q_i}{p_i} \right)}{\sum_{i=1}^{n} f(q_i)} \right) + \log_2 \frac{q}{p} .
\]

Setting now

\[
\log_2 \frac{q_i}{p_i} = x_i \quad \text{and} \quad \log_2 \frac{q}{p} = y
\]

and \( \psi(x + y) = \psi_y(x) \), we have

\[
\psi^{-1} \left( \frac{\sum_{i=1}^{n} f(p_i 2^{x_i}) \psi_y (x_i)}{\sum_{i=1}^{n} f(p_i 2^{x_i})} \right) = \psi^{-1} \left( \frac{\sum_{i=1}^{n} f(p_i 2^{x_i}) \psi (x_i)}{\sum_{i=1}^{n} f(p_i 2^{x_i})} \right)
\]

where

\[
\sum_{i=1}^{n} f(p_i 2^{x_i}) \leq 1.
\]
Now on all steps are similar. The two solutions in this case lead to (16) and (17).

This completes the proof of Theorem 2.

ACKNOWLEDGEMENT

My best thanks are due to late Professor R. S. Varma, D.Sc., F.N.I., F.N.A.Sc., ex-Dean of the Faculty of Mathematics, University of Delhi, under whose guidance this work was completed.

REFERENCES