KYBERNETIKA — VOLUME 11 (1975), NUMBER 4

On Additive and Non-Additive Entropies

D. P. MITTAL

Shannon's fundamental concept of entropy has been generalized in different directions. Rényi and Havrda-Charvát have defined 'entropies of order α'. Kerridge generalized the idea of entropy of a single probability distribution to a kind of cross-entropy of a pair of probability distributions called 'inaccuracy' such that Shannon's entropy is the minimum of Kerridge's inaccuracy. In this paper we have investigated functions, the minimum of one of which is Rényi's entropy of order α and that of the other is Vajda's entropy of order α.

1. INTRODUCTION

Let $(\Omega, \mathscr{B}, \mathscr{P})$ be a *probability space*; Ω is an abstract non-empty set consisting of elementary events x, $\mathscr{B}$ a σ-algebra of subsets of Ω containing Ω itself, $\mathscr{P}$ a *probability measure*, that is a non-negative countably additive set function defined on $\mathscr{B}$ such that $\mathscr{P}(\Omega) = 1$. Let

$$\Delta_n = \{(p_1, p_2, \ldots, p_n) = \mathscr{P}, \quad 0 < p_k \leqq 1, \quad 0 < \sum_{k=1}^{n} p_k \leqq 1\}, \quad n = 1, 2, 3, \ldots,$$

$$\Delta_n^* = \{(p_1, p_2, \ldots, p_n) = \mathscr{P}, \quad 0 < p_k \leqq 1, \quad \sum_{k=1}^{n} p_k = 1\}, \quad n = 1, 2, 3, \ldots,$$

denote the sets of n-components, $n \geqq 1$, generalized probability distributions and complete probability distributions respectively.

For $(p_1, p_2, \ldots, p_n) = \mathscr{P} \in \Delta_n$, $(q_1, q_2, \ldots, q_n) = \mathscr{Q} \in \Delta_n$, Nath [1] defined additive inaccuracy of order α as

$$(1.1) \qquad H_\alpha(\mathscr{P} \| \mathscr{Q}) = (1 - \alpha)^{-1} \log_2 \left(\sum_{k=1}^{n} p_k q_k^{\alpha-1} / \sum_{k=1}^{n} p_k\right), \quad \alpha > 0, \quad \alpha \neq 1,$$

$$= - \left(\sum_{k=1}^{n} p_k \log_2 q_k\right) / \sum_{k=1}^{n} p_k, \quad \alpha = 1,$$

and a non-additive inaccuracy of order α as

$$(1.2)\qquad h_\alpha(\mathscr{P} \parallel \mathscr{Q}) = (1 - 2^{1-\alpha})^{-1} \left[1 - \left(\sum_{k=1}^{n} p_k q_k^{\alpha-1}\right) \Big/ \sum_{k=1}^{n} p_k\right], \quad \alpha > 0, \quad \alpha \neq 1 .$$

If $\mathscr{P} \equiv \mathscr{Q}$, then $H_\alpha(\mathscr{P} \parallel \mathscr{Q})$ reduces to Rényi's [2] additive entropy $H_\alpha(\mathscr{P})$ of order α and $h_\alpha(\mathscr{P} \parallel \mathscr{Q})$ to non-additive entropy $h_\alpha(\mathscr{P})$ due to Vajda [3], where

$$(1.3)\qquad H_\alpha(\mathscr{P}) = (1 - \alpha)^{-1} \log_2 \left(\sum_{k=1}^{n} p_k^\alpha \Big/ \sum_{k=1}^{n} p_k\right), \quad \alpha > 0, \quad \alpha \neq 1 ,$$

$$= - \left(\sum_{k=1}^{n} p_k \log_2 p_k\right) \Big/ \sum_{k=1}^{n} p_k , \quad \alpha = 1 ,$$

and

$$(1.4)\qquad h_\alpha(\mathscr{P}) = (1 - 2^{1-\alpha})^{-1} \left[1 - \left(\sum_{k=1}^{n} p_k^\alpha \Big/ \sum_{k=1}^{n} p_k\right)\right], \quad \alpha > 0, \quad \alpha \neq 1 .$$

For $\mathscr{P} \in \Delta_n^*$ and $\alpha = 1$, $H_\alpha(\mathscr{P})$ reduces to Shannon's [5] entropy

$$(1.5)\qquad H_1(\mathscr{P}) = - \sum_{k=1}^{n} p_k \log_2 p_k$$

and $h_\alpha(\mathscr{P})$ reduces to Havrda-Charvát [6] entropy for $\alpha > 0$, $\alpha \neq 1$.

2. FORMULATION OF THE PROBLEM

If $\alpha = 1$ and $\mathscr{P} \in \Delta_n^*$, $\mathscr{Q} \in \Delta_n^*$, then an important property of Kerridge's inaccuracy [7] is that

$$(2.1)\qquad H_1(\mathscr{P}) \leqq H_1(\mathscr{P} \parallel \mathscr{Q}) ,$$

equality if and only if $\mathscr{P} \equiv \mathscr{Q}$. In other words, Shannon's entropy is the minimum value of Kerridge's inaccuracy. If $\mathscr{P} \in \Delta_n$, $\mathscr{Q} \in \Delta_n$, then (2.1) is no longer necessarily true. Also, the corresponding inequalities

$$(2.2)\qquad H_\alpha(\mathscr{P}) \leqq H_\alpha(\mathscr{P} \parallel \mathscr{Q}) , \quad \alpha > 0 , \quad \alpha \neq 1 ,$$

and

$$(2.3)\qquad h_\alpha(\mathscr{P}) \leqq h_\alpha(\mathscr{P} \parallel \mathscr{Q}) , \quad \alpha > 0 , \quad \alpha \neq 1 ,$$

are not necessarily true even for generalized probability distributions. Hence, it is natural to ask the following question:

"For generalized probability distributions, what are the quantities the minimum values of which are $H_\alpha(\mathscr{P})$ and $h_\alpha(\mathscr{P})$?" We give below an answer to the above question separately for $H_\alpha(\mathscr{P})$ and $h_\alpha(\mathscr{P})$ by dividing the discussion into two parts (i) $\alpha = 1$ and (ii) $\alpha \neq 1$, $\alpha > 0$. Also we shall assume that $n \geqq 2$, because the problem is trivial for $n = 1$.

3. RÉNYI'S ENTROPY

Case 1. Let $\alpha = 1$. If $\mathscr{P} \in \Delta_n^*$, $\mathscr{Q} \in \Delta_n^*$, then as remarked earlier (2.1) is true. For $\mathscr{P} \in \Delta_n$, $\mathscr{Q} \in \Delta_n$, it can be easily seen by using Jenson's inequality that (2.1) is true if $\sum_{k=1}^{n} p_k \geqq \sum_{k=1}^{n} q_k$, equality in (2.1) holding if and only if

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \ldots = \frac{p_n}{q_n} = \frac{\sum_{k=1}^{n} p_k}{\sum_{k=1}^{n} q_k}.$$

Case 2. Let $\alpha \neq 1$, $\alpha > 0$. Since (2.2) is not necessarily true, we need a function $F_\alpha(\mathscr{P}, \mathscr{Q})$ such that $H_\alpha(\mathscr{P}) \leqq F_\alpha(\mathscr{P}, \mathscr{Q})$, equality if and only if $\mathscr{P} \equiv \mathscr{Q}$.

Let

$$F_\alpha(\mathscr{P}, \mathscr{Q}) = (\alpha - 1)^{-1} \log_2 \left(\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha} \Big/ \sum_{k=1}^{n} p_k^\alpha \right). \tag{3.1}$$

By using Bellman's [4] principle of optimality, we shall show that $H_\alpha(\mathscr{P})$ is the minimum value of $F_\alpha(\mathscr{P}, \mathscr{Q})$ and the minimum value is attained when $\mathscr{P} \equiv \mathscr{Q}$.

In order to minimize $F_\alpha(\mathscr{P}, \mathscr{Q})$, it is enough to minimize $\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha}$ for $\alpha > 1$ and minimize $-\left(\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha}\right)$ for $0 < \alpha < 1$, under the constraints

$$p_1 + p_2 + \ldots + p_n = c_1 \leqq 1,$$

$$q_1 + q_2 + \ldots + q_n = c_2 \leqq 1.$$

Let

$$f_n(c_1, c_2) = \min \sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha}, \quad p_k > 0, \quad q_k > 0, \quad k = 1 \text{ to } n. \tag{3.2}$$

$$p_1 + p_2 + \ldots + p_n = c_1 \leqq 1,$$

$$q_1 + q_2 + \ldots + q_n = c_2 \leqq 1.$$

By using Bellman's principle of optimality, it follows that $f_n(c_1, c_2)$ satisfies the functional equation

$$f_n(c_1, c_2) = \min\left[x^\alpha y^{1-\alpha} + f_{n-1}(c_1 - x, c_2 - y)\right], \quad n \geqq 2,$$

$$0 < x \leqq c_1, \quad 0 < y \leqq c_2 .$$

Obviously,

$$f_1(c_1, c_2) = c_1^\alpha c_2^{1-\alpha},$$

$$f_2(c_1, c_2) = \min\left[x^\alpha y^{1-\alpha} + (c_1 - x)^\alpha (c_2 - y)^{1-\alpha}\right],$$

$$0 < x \leqq c_1, \quad 0 < y \leqq c_2 .$$

For extremal values,

$$\frac{\partial f_2}{\partial x} = 0 = \frac{\partial f_2}{\partial y} .$$

Actual computation gives $y = c_2/c_1 \,.\, x$. Also it can be easily verified that $\partial^2 f_2/\partial x^2 > 0$ and

$$\frac{\partial^2 f_2}{\partial x^2} \cdot \frac{\partial^2 f_2}{\partial y^2} - \left(\frac{\partial^2 f_2}{\partial x\, \partial y}\right)^2 > 0,$$

so that the condition $y = c_2/c_1 \,.\, x$, is a condition under which f_2 assumes its minimum value and the minimum value of $f_2(c_1, c_2)$ is $c_1^\alpha c_2^{1-\alpha}$. By following the above procedure, it can be shown that each of the functions $f_n(c_1, c_2)$ for $n \geqq 2$, achieves its minimum value when $y = c_2/c_1 \,.\, x$ and the minimum value is $c_1^\alpha c_2^{1-\alpha}$.

The case for $0 < \alpha < 1$ follows on the similar lines and the corresponding conclusion follows. Thus we conclude that $F_\alpha(\mathscr{P}, \mathscr{Q})$ achieves its minimum value, when

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \ldots = \frac{p_n}{q_n} = \frac{c_1}{c_2}$$

and the minimum value of $F_\alpha(\mathscr{P}, \mathscr{Q})$ is $H_\alpha(\mathscr{P})$. Consequently,

$$H_\alpha(\mathscr{P}) \leqq F_\alpha(\mathscr{P}, \mathscr{Q}), \quad \sum_{k=1}^{n} p_k \geqq \sum_{k=1}^{n} q_k, \tag{3.3}$$

equality being true iff the elements of $\mathscr{P}$ and $\mathscr{Q}$ are proportional. Note that (3.3) may be regarded as a generalization of Shannon's inequality.

Interpretation of $F_\alpha(\mathscr{P}, \mathscr{Q})$. The quantity $F_\alpha(\mathscr{P}, \mathscr{Q})$ can be interpreted as an inaccuracy of order α as follows:

By applying linear transformations as functions of α alone, several measures of inaccuracy can be derived from $H_\alpha(\mathscr{P} \parallel \mathscr{Q})$. For example, let us consider

$$\tilde{H}_\alpha(\mathscr{P} \parallel \mathscr{Q}) = (\alpha - 1)^{-1} \log_2 \left(\sum_{k=1}^{n} p_k q_k^{1-\alpha} / \sum_{k=1}^{n} p_k\right), \quad \alpha > 0, \quad \alpha \neq 1,$$

$$= H_1(\mathscr{P} \parallel \mathscr{Q}), \quad \alpha = 1.$$

Obviously,

$$\tilde{H}_\alpha(\mathscr{P} \parallel \mathscr{Q}) = H_{2-\alpha}(\mathscr{P} \parallel \mathscr{Q}), \quad 0 < \alpha < 2.$$

For $2 < \alpha < \infty$, $\tilde{H}_\alpha(\mathscr{P} \parallel \mathscr{Q})$ is defined independently.

Clearly, if $\mathscr{P}^{(\alpha)} = (p_1^{(\alpha)}, p_2^{(\alpha)}, \ldots, p_n^{(\alpha)})$,

where

$$p_k^{(\alpha)} = p_k^\alpha / \sum_{j=1}^{n} p_j^\alpha,$$

then for $\alpha \neq 1$,

$$\tilde{H}_\alpha(\mathscr{P}^{(\alpha)} \parallel \mathscr{Q}) = F_\alpha(\mathscr{P}, \mathscr{Q}).$$

4. VAJDA'S NON-ADDITIVE ENTROPY

Case 1. $\alpha = 1$. This is the same as Case 1 in Section 3.

Case 2. Let $\alpha \neq 1$, $\alpha > 0$. We *define* functions $\psi_\alpha : R \to R$ such that

$$\psi_\alpha(x) = (1 - 2^{1-\alpha})^{-1} \left[1 - 2^{(1-\alpha)x}\right], \quad \alpha \neq 1. \tag{4.1}$$

Obviously,

$$\psi_\alpha(H_\alpha(\mathscr{P})) = h_\alpha(\mathscr{P})$$

and

$$\psi_\alpha(\tilde{H}_\alpha(\mathscr{P}^{(\alpha)} \parallel \mathscr{Q})) = (1 - 2^{1-\alpha})^{-1} \left[1 - \left(\sum_{k=1}^{n} p_k^\alpha / \sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha}\right)\right], \quad \alpha \neq 1. \tag{4.2}$$

By using Bellman's principle of optimality, it can be established again that $\psi_\alpha(\tilde{H}_\alpha(\mathscr{P}^{(\alpha)} \parallel \mathscr{Q}))$ achieves its minimum value, when

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \ldots = \frac{p_n}{q_n} = \frac{\sum_{k=1}^{n} p_k}{\sum_{k=1}^{n} q_k}$$

and the minimum value is $h_\alpha(\mathscr{P})$, so that

$$h_\alpha(\mathscr{P}) \leqq \psi_\alpha(\tilde{H}_\alpha(P^{(\alpha)} \parallel \mathscr{Q})) = \psi_\alpha(F_\alpha(\mathscr{P}, \mathscr{Q}))\,.$$

This is another generalization of Shannon's inequality. Note that in this case, it is again enough to minimize $(\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha})$ for $\alpha > 1$ and minimize $(-\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha})$ for $0 < \alpha < 1$ under the conditions

$$p_1 + p_2 + \ldots + p_n = c_1 \leqq 1\,,$$
$$q_1 + q_2 + \ldots + q_n = c_2 \leqq 1\,,$$
$$p_k > 0\,,\quad q_k > 0\,,\quad k = 1 \text{ to } n\,.$$

I am thankful to Dr. Prem Nath for proposing the problem and his valuable suggestions. Thanks are also due to Dr. D. L. Jain for his kind discussions.

(Received May 29, 1973.)

REFERENCES

[1] P. Nath: An axiomatic characterization of inaccuracy for finite generalized discrete probability distributions. Opsearch *7* (1970), 115–133.

[2] A. Rényi: On measures of entropy and information. Proc. Fourth Berk. Symp. in Math. Stat. and Prob. 1 (1961), 547–561.

[3] I. Vajda: Axioms for α-entropy of a generalized probability scheme. Kybernetika *4* (1968), 2, 105–122.

[4] R. Bellman: Dynamic Programming. Princeton University Press, New Jersey 1957.

[5] C. E. Shannon: A mathematical theory of communication. B. S. T. J. *27* (1948), 379–423, 623–656.

[6] J. Havrda, F. Charvát: Quantification method of classification processes. The concept of structural α-entropy. Kybernetika *3* (1967), 1, 30–35.

[7] D. F. Kerridge: Inaccuracy and inference. Jour. Roy. Statist. Soc. Ser B *23* (1961), 184–196.

D. P. Mittal; Faculty of Mathematics, University of Delhi, Delhi — 7. India.