Available online at www.elixirpublishers.com (Elixir International Journal)

Applied Mathematics

Elixir Appl. Math. 39 (2011) 4616-4619

ISSN: 2229-712X

A generalization of Shannon inequality based on Aczel and Daroczy entropy and its application in coding theory

Satish Kumar and Arun Choudhary

Department of Mathematics, Geeta Institute of Management & Technology, Kanipla-136131, Kurukshetra, Haryana (India).

ARTICLE INFO

Article history:
Received: 21 July 2011;
Received in revised form:
19 September 2011;
Accepted: 26 September 2011;

Keywords
Renyi's Entropy,
Shannon inequality,
Codeword length,
Optimal code length,
Kraft inequality.

ABSTRACT

In the present paper, we have generalized Shannon inequality for Aczel and Daroczy entropy and give its application in coding theory. We define mean codeword length and their bounds have been defined and a coding theorem on a lower and upper bounds of a generalized mean codeword length in terms of Aczel and Daroczy entropy has been proved.

AMS Subject classification: 94A15, 94A17, 94A24, 26D15.

Introduction

Throughout the paper $\mathbb{N}$ denotes the set of the natural numbers and for $N \in \mathbb{N}$, we set

$$\Delta_N = \left\{ P = (p_1, p_2, ..., p_N); \, p_i \geq 0 \; and \sum_{i=1}^{N} p_i = 1 \right\}, \; N \geq 2$$

be the set of all finite discrete probability distributions. For any $(p_1, p_2, ..., p_N) = P \in \Delta_N$.

Shannon [12] defined entropy as:

$$H_1(P) = -\sum p_i \log p_i. \tag{1}$$

The measure (1) has been generalized by various authors and has found applications in various disciplines such as crime, economics, accounting, physics etc.

Throughout this paper, $\sum$ will stand for $\sum_{i=1}^{N}$ unless otherwise stated and logarithms are taken to the base $D(D > 1)$.

If $(q_1, q_2, ..., q_N) = Q \in \Delta_N$, is another probability distributions, it is well known that

$$-\sum p_i \log p_i \leq -\sum p_i \log q_i, \tag{2}$$

with equality iff $p_i = q_i, \forall i = 1,2,...,N$. The inequality (2) may be alternatively written in the form

$$H_1(P) \leq H_1(P,Q), \tag{3}$$

where

$$H_1(P,Q) = -\sum p_i \log q_i, \tag{4}$$

denotes Kerridge (1961) inaccuracy.

Nath (1970) defined inaccuracy of order α as:

$$H_\alpha(P,Q) = \frac{1}{1-\alpha} \log\left(\sum p_i q_i^{\alpha-1}\right), \quad \alpha > 0 (\neq 1). \tag{5}$$

The importance of (2) is well known in coding theory. The inequality (2) is commonly known as Shannon's (1948) inequality.

The objective of this paper is to study generalization of (2), and then give an applications in coding theory.

Generalization of Shannon's Inequality

Definition: Let $N \in \mathbb{N}$ be arbitrarily fixed, $\alpha, \beta > 0, \alpha \neq 1$ be given real numbers. Then the information measure $H_\alpha^\beta : \Delta_N \to \mathbb{R}$ is defined by

$$H_\alpha^\beta(P,Q) = \frac{1}{1-\alpha} \log\left(\frac{\sum p_i^\beta q_i^{(\alpha-1)}}{\sum p_i^\beta} \right) \tag{6}$$

$(p_1, p_2, ..., p_N) \in \Delta_N$; $(q_1, q_2, ..., q_N) \in \Delta_N$

Remarks

(i) If $\beta = 1$, then (6) reduces to (5).

(ii) If $\beta = 1, p_i = q_i$, then (6) reduces to Renyi's (1961) entropy.

$$i.e., \quad H_\alpha(P) = \frac{1}{1-\alpha} \log\left[\sum p_i^\alpha\right], \; \alpha > 0 (\neq 1). \tag{7}$$

(iii) If $\beta = 1$ and $\alpha \to 1$, then (6) reduces to Kerridge's (1961) inaccuracy.

(iv) If $\beta = 1, p_i = q_i$ and $\alpha \to 1$, then (6) reduces to (1).

(v) If $\alpha \to 1$, then (6) reduces to generalized Kerridge (1961) inaccuracy for the incomplete power distribution P^β as such:

$$i.e., \quad H^\beta(P,Q) = -\frac{\sum p_i^\beta \log q_i}{\sum p_i^\beta}, \; \beta > 0. \tag{8}$$

(vi) If $p_i = q_i$, then (6) becomes the entropy of order α and type β.

$$i.e., \quad H_\alpha^\beta(P) = \frac{1}{1-\alpha} \log\left(\frac{\sum p_i^{\alpha+\beta-1}}{\sum p_i^\beta} \right), \quad \alpha > 0 (\neq 1), \beta > 0. \tag{9}$$

Tele:
E-mail addresses: arunchoudhary07@gmail.com

Also, we call $H_\alpha^\beta(P)$is generalized Renyi's entropy of order α and type β studied by Aczel and Daroczy (1963) and Kapur (1967).

Since $H_\alpha^\beta(P,Q) \neq H_\alpha^\beta(P)$, we will not interpret (6) as a measure of inaccuracy. But $H_\alpha^\beta(P,Q)$ is a generalization of the measure of inaccuracy defined in (9). In spite of the fact that $H_\alpha^\beta(P,Q)$ is not a measure of inaccuracy in its usual sense, its study is justified because it leads to meaningful new measures of length. In the following theorem, we will determine a relation between (9) and (6) of the type (3).

Since (6) is not a measure of inaccuracy in its usual sense, we will call the generalized relation as pseudo-generalization of the Shannon inequality.

Now we are interested to extend the result of (2) in a fashion such as:

$$H_\alpha^\beta(P) \leq H_\alpha^\beta(P,Q). \tag{10}$$

Nath (1975) has shown that

$$H_\alpha(P) \leq H_\alpha(P,Q), \tag{11}$$

does not hold for all $\alpha > 0$, he showed that

$$H_\alpha(P) \leq H_\alpha(R_\alpha(P,Q),Q),\ \alpha > 0, \tag{12}$$

with equality iff $p_i = q_i, \forall i = 1,2,...,N,$

where

$$R_\alpha(P,Q) = (r_1(\alpha), r_2(\alpha)..., r_N(\alpha)), \tag{13}$$

and

$$r_i = \frac{p_i^\alpha q_i^{1-\alpha}}{\sum p_i^\alpha q_i^{1-\alpha}}. \tag{14}$$

Presently, we are interested to extend the above result as follows:

$$H_\alpha^\beta(P) \leq H_\alpha\left(R_\alpha^\beta(P,Q),Q\right), \tag{15}$$

where

$$H_\alpha^\beta(P) = \frac{1}{1-\alpha}\log\left(\frac{\sum p_i^{\alpha+\beta-1}}{\sum p_i^\beta}\right), \quad \alpha > 0(\neq 1), \beta > 0,$$

and

$$H_\alpha^\beta(P,Q) = \frac{1}{1-\alpha}\log\left(\frac{\sum p_i^\beta q_i^{(\alpha-1)}}{\sum p_i^\beta}\right), \quad \alpha > 0(\neq 1), \beta > 0.$$

Lemma 1: Let $(p_1, p_2,..., p_N) = P \in \Delta_N$, $(q_1, q_2,..., q_N) = Q \in \Delta_N,\ N \geq 2.$

And $R_\alpha^\beta(P,Q) = (r_1(\alpha,\beta), r_2(\alpha,\beta),..., r_N(\alpha,\beta)),$

where $r_i(\alpha,\beta) = \dfrac{p_i^{\alpha+\beta-1} q_i^{1-\alpha}}{\sum p_i^{\alpha+\beta-1} q_i^{1-\alpha}}$, i=1,2,...,N.

Then,

$$H_\alpha^\beta(P) \leq H_\alpha\left(R_\alpha^\beta(P,Q),Q\right), \tag{16}$$

with equality iff $\mathrm{p_i = q_i, \forall i = 1, 2,..., N}.$

Proof: If $\alpha = 1,$ then (16) becomes $-\sum p_i^\beta \log p_i \leq -\sum p_i^\beta \log q_i.$ For $\beta = 1, \alpha = 1,$ (16) reduces to (2). Here we give proof only for $\alpha > 0(\neq 1)$ and $\beta > 0$. It is easily seen that for $\alpha > 0,$ the probability distribution $R_\alpha^\beta(P,Q)$ are complete and belong to Δ_N. Simple computation gives,

$$H_\alpha\left(R_\alpha^\beta(P,Q),Q\right) = H_\alpha^\beta(P) + I_\alpha^\beta(P//Q), \tag{17}$$

where

$$\mathrm{I}_\alpha^\beta(\mathrm{P//Q}) = \frac{1}{\alpha-1}\log\left(\frac{\sum \mathrm{p_i}^{\alpha+\beta-1}\mathrm{q_i}^{(1-\alpha)}}{\sum \mathrm{p_i}^\beta}\right),\ \alpha > 0(\neq 1) \tag{18}$$

denotes Kapur (1968) relative information of order α and type β.

Since, $I_\alpha^\beta(P//Q) \geq 0$ for $q_i \leq p_i$,

and

$$I_\alpha^\beta(P//Q) = 0 \ \ iff \ \ p_i = q_i, \forall i = 1,2,..., N. \tag{19}$$

The inequality (16) follows immediately from (17) and (18).

From (16), it follows that

$$H_\alpha^\beta(P) = \inf_Q H_\alpha\left(R_\alpha^\beta(P,Q),Q\right). \tag{20}$$

Nath (1975) has defined Renyi's entropy as

$$H_\alpha(P) = \inf_{Q\in\Delta_n} H_\alpha(P,Q). \tag{21}$$

Obviously (20) is generalization of (21).

Now we explain utility of (2) and (15) in coding theory.

Let there be an ensemble of messages $x_1, x_2,..., x_N$ with probability of x_i being $p_i > 0, \sum p_i = 1.$ Suppose that the above messages are encoded in uniquely decipherable way by using letters from an alphabet $A = (a_1, a_2,..., a_D)$ called the code alphabet. If the length of code word assigned to message x_i is n_i, then Mc-Millan (1956) inequality

$$\sum D^{-n_i} \leq 1, \tag{22}$$

holds. Also the traditional average length

$$L = \sum n_i p_i \tag{23}$$

satisfies the inequality

$$L \geq H_1(P), \tag{24}$$

with equality iff $p_i = D^{-n_i}, \forall i = 1,2,3,..., N.$

The classical noiseless coding theorem states that

$$H(P) \leq L < H(P) + 1. \tag{25}$$

Campbell (1965) has defined the average codeword length of order t as

$$L_t = \frac{1}{t}\log\left[\sum p_i D^{tn_i}\right],\ -1 < t < \infty, \tag{26}$$

and proved that if

$$\alpha = (1+t)^{-1} \tag{27}$$

then,

$$L_{t(\alpha)} \geq H_\alpha(P), \tag{28}$$

with equality iff

$$n_i = -\alpha \log p_i + \log\left(\sum p_i^{\alpha}\right). \qquad (29)$$

Further, Nath (1975) also extended the noiseless coding theorem (26) in the form,

$$H_{\alpha}(P) \le L_{\alpha} < H_{\alpha}(P)+1, \qquad (30)$$

where

$$L_{\alpha} = \frac{1}{\alpha-1}\log\left[\frac{\sum p_i^{\alpha} D^{(\alpha-1)n_i}}{\sum p_i^{\alpha}}\right], \alpha > 0(\ne 1). \qquad (31)$$

Definition: Let $N \in \mathsf{N}$, $\alpha, \beta > 0$, $\alpha \ne 1$ be arbitrarily fixed, then the average code word length of order α and type β corresponding to the generalized information measure $H_{\alpha}^{\beta}(P,Q)$ is given by the formula

$$L(\alpha,\beta) = \frac{1}{\alpha-1}\log\left(\frac{\sum p_i^{\alpha+\beta-1} D^{(\alpha-1)n_i}}{\sum p_i^{\alpha+\beta-1}}\right). \qquad (32)$$

Remarks

(i) When $\beta = 1$, then (32) reduces to average code word length studied by Nath (1975).

(ii) When $\beta = 1$ and $\alpha \to 1$, then (32) reduces to average codeword length studied by Shannon (1948).

Measure of average codeword length of order α and type β

Definition: Let t_i denote the cost of transmitting the letter a_i of the code alphabet A. If $t_i = 1, i = 1,2,...,D,$ then the average cost of transmission per massage is nothing but L given by (23). Let us define the average code length

$$Z(P, <n_i>, f, \phi) = \phi^{-1}\left[\frac{\sum f(p_i)\phi(n_i)}{\sum f(p_i)}\right]. \qquad (33)$$

Where f is a non-constant positive valued, continuous function defined on (0,1], f is a strictly monotonically increasing and continuous real valued function defined on $[1,\infty)$ such that ϕ^{-1} exists. From (33), it is clear that we are defining average code length as a most gerenal quasilinear mean value rather than ordinary average.

An intuitive requirement which any measure of average code length should satisfy is that it should be translative for all positive integers $m \in N^{+}$, Where N^{+} denote the set of all positive integers. In other word, if the length of each code word is increased by a positive integer $m \in N^{+}$ by attaching to the right, sequence of length m constructed by using letter of code alphabet A, then the average code length must be increased by m. Thus, we get the functional equation

$$\phi^{-1}\left[\frac{\sum f(p_i)\phi(n_i+m)}{\sum f(p_i)}\right] = \phi^{-1}\left[\frac{\sum f(p_i)\phi(n_i)}{\sum f(p_i)}\right] + m, \forall\, m \in N^{+}. \qquad (34)$$

Equation (34) is known as translative equation following J. Aczel (1974), the following theorem can be proved easily.

Theorem 1: The only quasilinear measure $Z(P, <n_i>, f, \phi)$ of average code length which are translative $\forall m \in N^{+}$ are

$$(a)\; Z(P, <n_i>, f, \phi_1) = \frac{\sum f(p_i) n_i}{\sum f(p_i)}. \qquad (35)$$

$$\text{(b)}\; Z(P, <n_i>, f, \phi_{\alpha}) = \frac{1}{\alpha-1}\log\left[\frac{\sum f(p_i) D^{(\alpha-1)n_i}}{\sum f(p_i)}\right],\; \alpha > 0(\ne 1) \qquad (36)$$

Where $\phi_1(x) = bx + a, a \ne 0, b > 0$ and $\phi_{\alpha}(x) = bD^{(\alpha-1)x} + a;\; D > 1, b > 0$.

Now, our object is to connect (36) to entropy of order α and type β. Let us write (36) in the form:

$$Z(P, <n_i>, f, \phi_{\alpha}) = \frac{1}{\alpha-1}\log\left[\sum f(p_i) D^{(\alpha-1)n_i}\right] + \frac{1}{1-\alpha}\log\left[\sum f(p_i)\right]. \qquad (37)$$

We restrict to $\alpha > 0$ and choose $f(p_i)$ in such a way that the last term on the R.H.S. in (37) becomes (9).

$$\text{i.e.,} \frac{1}{1-\alpha}\log\left[\sum f(p_i)\right] = \frac{1}{1-\alpha}\log\left(\frac{\sum p_i^{\alpha+\beta-1}}{\sum p_i^{\beta}}\right),\; \alpha > 0(\ne 1) \qquad (38)$$

Equation (38) gives rise to the functional equation

$$\sum f(p_i) = \frac{\sum p_i^{\alpha+\beta-1}}{\sum p_i^{\beta}}, \alpha > 0(\ne 1), \beta > 0,\; n = 2,3,... . \qquad (39)$$

The only non-constant continuous solution of (39) (refer also to Aczel, (1966)) are of the form

$$f(p_i) = \frac{p_i^{\alpha+\beta-1}}{\sum p_i^{\beta}}, \alpha > 0(\ne 1), \beta > 0, 0 < p_i \le 1, \forall m \in N^{+}. \qquad (40)$$

Hence,

$$Z(P, <n_i>, f, \phi_{\alpha}) = L(\alpha,\beta), \qquad (41)$$

where

$$L(\alpha,\beta) = \frac{1}{\alpha-1}\log\left(\frac{\sum p_i^{\alpha+\beta-1} D^{(\alpha-1)n_i}}{\sum p_i^{\alpha+\beta-1}}\right),\; \alpha > 0(\ne 1) \qquad (42)$$

In the following theorem, we give a relation between $L(\alpha,\beta)$ and $H_{\alpha}^{\beta}(P)$.

Theorem 2: If $n_i, i = 1,2,..., N$ are the lengths of codewords satisfying (22), then

$$H_{\alpha}^{\beta}(P) \le L(\alpha,\beta) < H_{\alpha}^{\beta}(P)+1,\; \beta > 0. \qquad (43)$$

The sign of equality holds iff

$$p_i = D^{-n_i}. \qquad (44)$$

Proof : In (15) choose $Q = (q_1, q_2, ..., q_N)$ where

$$q_i = D^{-n_i} \qquad (45)$$

with choice of Q, (15) becomes

$$H_{\alpha}^{\beta}(P) \le \frac{1}{\alpha-1}\log\left(\frac{\sum p_i^{\alpha+\beta-1} D^{(\alpha-1)n_i}}{\sum p_i^{\alpha+\beta-1}}\right).$$

i.e., $H_{\alpha}^{\beta}(P) \le L(\alpha,\beta)$ which proves the first part of (43).

The equality holds iff $D^{-n_i} = p_i,\ i = 1,2,...,N$ which is equivalent to

$$n_i = -\log_D p_i. \tag{46}$$

Choose all n_i such that

$$\log\frac{1}{p_i} \le n_i < \log\frac{1}{p_i} + 1. \tag{47}$$

Using the above relation, it follows that

$$D^{-n_i} > D^{-1} p_i. \tag{48}$$

We now have two possibilities:

If $\alpha > 1$; (48) gives us

$$p_i^{\alpha+\beta-1} D^{(\alpha-1)n_i} < p_i^{\beta} D^{(\alpha-1)}, \tag{49}$$

multiplying both side by $\dfrac{1}{\sum p_i^{\alpha+\beta-1}}$ and then summing over i, we get the required result.

2) If $0 < \alpha < 1$. The proof follows on the same lines.

References

Aczel, J., Lectures on Functional Equation and their Applications, Academic Press, New York, 1966.

Aczel, J., Determination of all additive quasiarithmetic mean codeword lengths, Z. Wahr. Verw. Geb., vol. 29, 1974, pp. 351-360.

Aczel, J. and Z. Daroczy, Uber Verallegemeineste quasiliniare mittelveste die mit grewinebts functionen gebildet Sind. Pub. Math. Debrecan, vol. 10, 1963, pp. 171-190.

Campbell, L.L., A coding theorem and Renyi's entropy, Information and Control, vol. 8, 1965, pp. 423-429.

Kapur, J.N., Generalized entropy of order α and type β, Maths. Seminar, Delhi, vol. 4, 1967, pp. 78-94.

Kapur, J.N., Information of order α and type β, Proc. Indian Acad. Sci., A-68, 1968, pp. 65-75.

Kerridge, D.F., Inaccuracy and inference, J.R. Stat. Soc., Ser. B, vol. 23, 1961, 184-194.

Mc-Millan, B., Two inequalities implied by using decipherability, IEEE Trans. Inform. Theory, vol. 2, no. 4, 1956, 115-116.

Nath, P., An axiomatic characterization of inaccuracy for discrete generalized probability distribution, Opsearch, vol. 7, 1970, pp. 115-133.

Nath, P., On a Coding Theorem Connected with Renyi's Entropy, Information and Control, vol. 29, 1975, pp. 234-242.

Renyi, A., On Measure of entropy and information, Proc. 4th Berkeley Symp. Maths. Stat. Prob., vol. 1, 1961, pp. 547-561.

Shannon, C.E., A Mathematical Theory of Communication, Bell System Tech. J., vol. 27, 1948, pp. 379-423, 623-656.