

## Research Article

# Some Inequalities in Information Theory Using Tsallis Entropy

**Litegebe Wondie and Satish Kumar** 

Department of Mathematics, College of Natural and Computational Science, University of Gondar, Gondar, Ethiopia

Correspondence should be addressed to Satish Kumar; drsatish74@rediffmail.com

Received 27 December 2017; Accepted 27 February 2018; Published 3 April 2018

Academic Editor: A. Zayed

Copyright © 2018 Litegebe Wondie and Satish Kumar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a relation between Tsallis's entropy and generalized Kerridge inaccuracy which is called generalized Shannon inequality and is well-known generalization in information theory and then give its application in coding theory. The objective of the paper is to establish a result on noiseless coding theorem for the proposed mean code length in terms of generalized information measure of order  $\xi$ .

## 1. Introduction

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

$$\Delta_n = \left\{ A = (a_1, a_2, \dots, a_n); 0 < a_k \leq 1, 0 < \sum_{k=1}^n a_k \leq 1 \right\}, \quad (1)$$

$$\Delta_n^* = \left\{ A = (a_1, a_2, \dots, a_n); 0 < a_k \leq 1, \sum_{k=1}^n a_k = 1 \right\},$$

where  $n = 2, 3, 4, \dots$  denote the set of all  $n$ -components complete and incomplete discrete probability distributions.

For  $(a_1, a_2, \dots, a_n) = A \in \Delta_n$ ,  $(b_1, b_2, \dots, b_n) = B \in \Delta_n$ , we define a nonadditive measure of inaccuracy, denoted by  $H(A, B; \xi)$  as

$$H(A, B; \xi) = \frac{1}{1 - \xi} \left[ \left( \frac{\sum_{k=1}^n a_k^{(\xi^2 - \xi + 1)/\xi} b_k^{((\xi - 1)/\xi)}}{(\sum_{k=1}^n a_k)^{1/\xi}} \right)^\xi - 1 \right] \quad (2)$$

$\xi > 1$

$$= -\frac{\sum_{k=1}^n a_k \log_2 b_k}{\sum_{k=1}^n a_k}; \quad \xi \rightarrow 1.$$

If  $b_k = a_k / \sum_{k=1}^n a_k^\xi$ , then  $H(A, B; \xi)$  reduces to nonadditive entropy.

$$\text{i.e., } H(A; \xi) = \frac{1}{1 - \xi} \left[ \left( \frac{\sum_{k=1}^n a_k^\xi}{\sum_{k=1}^n a_k} \right) - 1 \right] \quad \xi > 0 (\neq 1) \quad (3)$$

$$= -\frac{\sum_{k=1}^n a_k \log_2 a_k}{\sum_{k=1}^n a_k}; \quad \xi \rightarrow 1.$$

Entropy (3) was first of all characterized by Havrda and Charvát [1]. Later on, Daróczy [2] and Behara and Nath [3] studied this entropy. Vajda [4] also characterized this entropy for finite discrete generalized probability distributions. Sharma and Mittal [5] generalized this measure which is known as the entropy of order  $\alpha$  and type  $\beta$ . Pereira and Gur Dial [6] and Gur Dial [7] also studied Sharma and Mittal entropy for a generalization of Shannon inequality and gave its applications in coding theory. Kumar and Choudhary [8] also gave its application in coding theory. Recently, Wondie and Kumar [9] gave a Joint Representation of Renyi's and Tsallis Entropy. Tsallis [10] gave its applications in physics for  $A \in \Delta_n^*$ , and  $\xi \rightarrow 1$ ,  $H(A; \xi)$  reduces to Shannon [11] entropy

$$\text{i.e., } H(A) = -\sum_{k=1}^n a_k \log_D a_k. \quad (4)$$

Inequality (6) has been generalized in the case of Renyi's entropy.

### 2. Formulation of the Problem

For  $\xi \rightarrow 1$  and  $A \in \Delta_n^*, B \in \Delta_n^*$ , then an important property of Kerridge's inaccuracy [12] is that

$$H(A) \leq H(A, B). \tag{5}$$

Equality holds if and only if  $A = B$ . In other words, Shannon's entropy is the minimum value of Kerridge's inaccuracy. If  $A \in \Delta_n, B \in \Delta_n$ , then (5) is no longer necessarily true. Also, the corresponding inequality

$$H(A; \xi) \leq H(A, B; \xi) \tag{6}$$

is not necessarily true even for generalized probability distributions. Hence, it is natural to ask the following question: "For generalized probability distributions, what are the quantity the minimum values of which are  $H(A; \xi)$ ?" We give below an answer to the above question separately for  $H(A; \xi)$  by dividing the discussion into two parts (i)  $\xi \rightarrow 1$  and (ii)  $1 \neq \xi$ . Also we shall assume that  $n \geq 2$ , because the problem is trivial for  $n = 1$ .

*Case 1.* Let  $\xi \rightarrow 1$ . If  $A \in \Delta_n^*, B \in \Delta_n^*$ , then as remarked earlier (5) is true. For  $A \in \Delta_n, B \in \Delta_n$ , it can be easily seen by using Jenson's inequality that (5) is true if  $\sum_{k=1}^n a_k \geq \sum_{k=1}^n b_k$ , equality in (5) holding if and only if

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \dots = \frac{a_n}{b_n} = \frac{\sum_{k=1}^n a_k}{\sum_{k=1}^n b_k}. \tag{7}$$

*Case 2.* Let  $1 \neq \xi$ . Since (6) is not necessarily true, we need an inequality

$$\sum_{k=1}^n a_k^{\xi-1} b_k \leq 1; \quad \xi > 1 \tag{8}$$

such that  $H(A; \xi) \leq H(A, B; \xi)$  and equality holds if and only if  $b_k = a_k / \sum_{k=1}^n a_k^{\xi}$ .

Since  $\xi > 1$ , by reverse Hölder inequality, that is, if  $n = 2, 3, \dots, \gamma > 1$  and  $x_1, \dots, x_n, y_1, \dots, y_n$  are positive real numbers, then

$$\left( \sum_{k=1}^n x_k^{1/\gamma} \right)^\gamma \cdot \left( \sum_{k=1}^n y_k^{-1/(\gamma-1)} \right)^{-(\gamma-1)} \leq \sum_{k=1}^n x_k y_k. \tag{9}$$

Let  $\gamma = \xi/(\xi - 1)$ ,  $x_k = a_k^{(\xi^2 - \xi + 1)/(\xi - 1)} b_k$ , and  $y_k = a_k^{\xi/(1 - \xi)}$  ( $k = 1, 2, 3, \dots, n$ ).

Putting these values into (9), we get

$$\begin{aligned} & \left( \sum_{k=1}^n a_k^{(\xi^2 - \xi + 1)/\xi} b_k^{(\xi - 1)/\xi} \right)^{\xi/(\xi - 1)} \left( \sum_{k=1}^n a_k^{\xi} \right)^{1/(1 - \xi)} \\ & \leq \sum_{k=1}^n a_k^{\xi - 1} b_k \leq 1, \end{aligned} \tag{10}$$

where we used (8), too. This implies however that

$$\left( \sum_{k=1}^n a_k^{(\xi^2 - \xi + 1)/\xi} b_k^{(\xi - 1)/\xi} \right)^{\xi} \leq \left( \sum_{k=1}^n a_k^{\xi} \right). \tag{11}$$

Or

$$\left( \frac{\sum_{k=1}^n a_k^{(\xi^2 - \xi + 1)/\xi} b_k^{(\xi - 1)/\xi}}{(\sum_{k=1}^n a_k)^{1/\xi}} \right)^{\xi} \leq \left( \frac{\sum_{k=1}^n a_k^{\xi}}{\sum_{k=1}^n a_k} \right); \tag{12}$$

using (12) and the fact that  $\xi > 1$ , we get (6).

*Particular's Case.* If  $\xi = 1$ , then (6) becomes

$$H(A) \leq H(A, B), \tag{13}$$

which is Kerridge's inaccuracy [12].

### 3. Mean Codeword Length and Its Bounds

We will now give an application of inequality (6) in coding theory for

$$\Delta_n^* = \left\{ A = (a_1, a_2, \dots, a_n); 0 < a_k \leq 1, \sum_{k=1}^n a_k = 1 \right\}. \tag{14}$$

Let a finite set of  $n$  input symbols

$$X = \{x_1, x_2, \dots, x_n\} \tag{15}$$

be encoded using alphabet of  $D$  symbols, then it has been shown by Feinstein [13] that there is a uniquely decipherable code with lengths  $N_1, N_2, \dots, N_n$  if and only if the Kraft inequality holds; that is,

$$\sum_{k=1}^n D^{-N_k} \leq 1, \tag{16}$$

where  $D$  is the size of code alphabet.

Furthermore, if

$$L = \sum_{k=1}^n N_k a_k \tag{17}$$

is the average codeword length, then for a code satisfying (16), the inequality

$$L \geq H(A) \tag{18}$$

is also fulfilled and equality holds if and only if

$$N_k = -\log_D(a_k) \quad (k = 1, \dots, n), \tag{19}$$

and by suitable encoded into words of long sequences, the average length can be made arbitrarily close to  $H(A)$ , (see Feinstein [13]). This is Shannon's noiseless coding theorem.

By considering Renyi's entropy (see, e.g., [14]), a coding theorem, analogous to the above noiseless coding theorem, has been established by Campbell [15] and the authors obtained bounds for it in terms of

$$H_\xi(A) = \frac{1}{1 - \xi} \log_D \sum_{k=1}^n a_k^\xi, \quad \xi > 0 (\neq 1). \tag{20}$$

Kieffer [16] defined a class rules and showed that  $H_\xi(A)$  is the best decision rule for deciding which of the two sources can be coded with expected cost of sequences of length  $N$  when  $N \rightarrow \infty$ , where the cost of encoding a sequence is assumed to be a function of length only. Further, in Jelinek [17] it is shown that coding with respect to Campbell's mean length is useful in minimizing the problem of buffer overflow which occurs when the source symbol is produced at a fixed rate and the code words are stored temporarily in a finite buffer. Concerning Campbell's mean length the reader can consult [15].

It may be seen that the mean codeword length (17) had been generalized parametrically by Campbell [15] and their bounds had been studied in terms of generalized measures of entropies. Here we give another generalization of (17) and study its bounds in terms of generalized entropy of order  $\xi$ .

Generalized coding theorems by considering different information measure under the condition of unique decipherability were investigated by several authors; see, for instance, the papers [6, 13, 18].

An investigation is carried out concerning discrete memoryless sources possessing an additional parameter  $\xi$  which seems to be significant in problem of storage and transmission (see [9, 16–18]).

In this section we study a coding theorem by considering a new information measure depending on a parameter. Our motivation is, among others, that this quantity generalizes some information measures already existing in the literature such as Arndt [19] entropy, which is used in physics.

*Definition 1.* Let  $n \in \mathbb{N}$ ,  $\xi > 0 (\neq 1)$  be arbitrarily fixed, then the mean length  $L(\xi)$  corresponding to the generalized information measure  $H(A; \xi)$  is given by the formula

$$L(\xi) = \frac{1}{1-\xi} \left[ \left( \sum_{k=1}^n a_k^{(\xi^2-\xi+1)/\xi} D^{N_k((1-\xi)/\xi)} \right)^\xi - 1 \right], \quad (21)$$

where  $A = (a_1, \dots, a_n) \in \Delta_n^*$  and  $D, N_1, N_2, \dots, N_n$  are positive integers so that

$$\sum_{k=1}^n a_k^{\xi-1} D^{-N_k} \leq 1. \quad (22)$$

Since (22) reduces to Kraft inequality when  $\xi = 1$ , therefore it is called generalized Kraft inequality and codes obtained under this generalized inequality are called personal codes.

**Theorem 2.** Let  $n \in \mathbb{N}$ ,  $\xi > 1$  be arbitrarily fixed. Then there exist code length  $N_1, \dots, N_n$  so that

$$H(A; \xi) \leq L(\xi) < D^{1-\xi} H(A; \xi) + \frac{-1 + D^{1-\xi}}{1-\xi} \quad (23)$$

holds under condition (22) and equality holds if and only if

$$N_k = -\log_D \left( \frac{a_k}{\sum_{k=1}^n a_k^\xi} \right); \quad k = 1, 2, \dots, n, \quad (24)$$

where  $H(A; \xi)$  and  $L(\xi)$  are given by (3) and (21), respectively.

*Proof.* First of all we shall prove the lower bound of  $L(\xi)$ .

By reverse Hölder inequality, that is, if  $n = 2, 3, \dots, \gamma > 1$  and  $x_1, \dots, x_n, y_1, \dots, y_n$  are positive real numbers then

$$\left( \sum_{k=1}^n x_k^{1/\gamma} \right)^\gamma \cdot \left( \sum_{k=1}^n y_k^{-1/(\gamma-1)} \right)^{-(\gamma-1)} \leq \sum_{k=1}^n x_k y_k. \quad (25)$$

Let  $\gamma = \xi/(\xi-1)$ ,  $x_k = a_k^{(\xi^2-\xi+1)/(\xi-1)} D^{-N_k}$ , and  $y_k = a_k^{\xi/(1-\xi)}$  ( $k = 1, 2, 3, \dots, n$ ).

Putting these values into (25), we get

$$\begin{aligned} & \left( \sum_{k=1}^n a_k^{(\xi^2-\xi+1)/\xi} D^{-N_k((\xi-1)/\xi)} \right)^{\xi/(\xi-1)} \left( \sum_{k=1}^n a_k^\xi \right)^{1/(1-\xi)} \\ & \leq \sum_{k=1}^n a_k^{\xi-1} D^{-N_k} \leq 1; \quad \xi > 1, \end{aligned} \quad (26)$$

where we used (22), too. This implies however that

$$\left( \sum_{k=1}^n a_k^{(\xi^2-\xi+1)/\xi} D^{-N_k((\xi-1)/\xi)} \right)^{\xi/(\xi-1)} \leq \left( \sum_{k=1}^n a_k^\xi \right)^{1/(\xi-1)}. \quad (27)$$

For  $\xi > 1$ , (27) becomes

$$\left( \sum_{k=1}^n a_k^{(\xi^2-\xi+1)/\xi} D^{-N_k((\xi-1)/\xi)} \right)^\xi \leq \left( \sum_{k=1}^n a_k^\xi \right); \quad (28)$$

using (28) and the fact that  $\xi > 1$ , we get

$$H(A; \xi) \leq L(\xi). \quad (29)$$

From (24) and after simplification, we get

$$a_k^{(\xi^2-\xi+1)/\xi} D^{-N_k((\xi-1)/\xi)} = \frac{a_k^\xi}{\left( \sum_{k=1}^n a_k^\xi \right)^{(\xi-1)/\xi}}. \quad (30)$$

This implies

$$\left( \sum_{k=1}^n a_k^{(\xi^2-\xi+1)/\xi} D^{-N_k((\xi-1)/\xi)} \right)^\xi = \left( \sum_{k=1}^n a_k^\xi \right), \quad (31)$$

which gives  $L(\xi) = H(A; \xi)$ . Then equality sign holds in (29).

Now we will prove inequality (23) for upper bound of  $L(\xi)$ .

We choose the codeword lengths  $N_k, k = 1, \dots, n$  in such a way that

$$-\log_D \left( \frac{a_k}{\sum_{k=1}^n a_k^\xi} \right) \leq N_k < -\log_D \left( \frac{a_k}{\sum_{k=1}^n a_k^\xi} \right) + 1 \quad (32)$$

is fulfilled for all  $k = 1, \dots, n$ .

From the left inequality of (32), we have

$$D^{-N_k} \leq \frac{a_k}{\sum_{k=1}^n a_k^\xi}; \quad (33)$$

multiplying both sides by  $a_k^{\xi-1}$  and then taking sum over  $k$ , we get the generalized inequality (22). So there exists a generalized code with code lengths  $N_k$ ,  $k = 1, \dots, n$ .

Since  $\xi > 1$ , then (32) can be written as

$$\begin{aligned} \left( \frac{a_k}{\sum_{k=1}^n a_k^\xi} \right)^{(\xi-1)/\xi} &\geq D^{-N_k((\xi-1)/\xi)} \\ &> \left( \frac{a_k}{\sum_{k=1}^n a_k^\xi} \right)^{(\xi-1)/\xi} D^{-((\xi-1)/\xi)}. \end{aligned} \quad (34)$$

Multiplying (34) throughout by  $a_k^{(\xi^2-\xi+1)/\xi}$  and then summing up from  $k = 1$  to  $n$ , we obtain inequality

$$\begin{aligned} \left( \sum_{k=1}^n a_k^\xi \right) &\geq \left( \sum_{k=1}^n a_k^{(\xi^2-\xi+1)/\xi} D^{N_k((1-\xi)/\xi)} \right)^\xi \\ &> \left( \sum_{k=1}^n a_k^\xi \right) D^{(1-\xi)}. \end{aligned} \quad (35)$$

Since  $1 - \xi < 0$  for  $\xi > 1$ , we get from (35) inequality (23).  $\square$

*Particular's Cases.* For  $\xi \rightarrow 1$ , then (23) becomes

$$\frac{H(A)}{\log D} \leq L < \frac{H(A)}{\log D} + 1, \quad (36)$$

which is the Shannon [11] classical noiseless coding theorem

$$N_k = -\log_D(a_k); \quad k = 1, 2, \dots, n, \quad (37)$$

where  $H(A)$  and  $L$  are given by (4) and (17), respectively.

## 4. Conclusion

In this paper we prove a generalization of Shannon's inequality for the case of entropy of order  $\xi$  with the help of Hölder inequality. Noiseless coding theorem is proved. Considering Theorem 2 we remark that the optimal code length depends on  $\xi$  in contrast with the optimal code lengths of Shannon which do not depend of a parameter. However, it is possible to prove coding theorem with respect to (3) such that the optimal code lengths are identical to those of Shannon.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] J. Havrda and F. S. Charvát, "Quantification method of classification processes. Concept of structural  $\alpha$ -entropy," *Kybernetika*, vol. 3, pp. 30–35, 1967.
- [2] Z. Daróczy, "Generalized information functions," *Information and Control*, vol. 16, no. 1, pp. 36–51, 1970.
- [3] M. Behara and P. Nath, "Additive and non-additive entropies of finite measurable partitions," in *Probability and Information Theory II*, pp. 102–138, Springer-Verlag, 1970.
- [4] I. Vajda, "Axioms for  $\alpha$ -entropy of a generalized probability scheme," *Kybernetika*, vol. 4, pp. 105–112, 1968.
- [5] B. D. Sharma and D. P. Mittal, "New nonadditive measures of entropy for discrete probability distributions," *Journal of Mathematical Sciences*, vol. 10, pp. 28–40, 1975.
- [6] R. Pereira and Gur Dial, "Pseudogeneralization of Shannon inequality for Mittal's entropy and its application in coding theory," *Kybernetika*, vol. 20, no. 1, pp. 73–77, 1984.
- [7] Gur Dial, "On a coding theorems connected with entropy of order  $\alpha$  and type  $\beta$ ," *Information Sciences*, vol. 30, no. 1, pp. 55–65, 1983.
- [8] S. Kumar and A. Choudhary, "Some coding theorems on generalized Havrda-Charvat and Tsallis's entropy," *Tamkang Journal of Mathematics*, vol. 43, no. 3, pp. 437–444, 2012.
- [9] L. Wondie and S. Kumar, "A joint representation of Renyi's and Tsalli's entropy with application in coding theory," *International Journal of Mathematics and Mathematical Sciences*, vol. 2017, Article ID 2683293, 5 pages, 2017.
- [10] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.
- [11] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.
- [12] D. F. Kerridge, "Inaccuracy and inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 23, pp. 184–194, 1961.
- [13] A. Feinstein, *Foundations of Information Theory*, McGraw-Hill, New York, NY, USA, 1956.
- [14] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561, University of California Press, 1961.
- [15] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965.
- [16] J. C. Kieffer, "Variable-length source coding with a cost depending only on the code word length," *Information and Control*, vol. 41, no. 2, pp. 136–146, 1979.
- [17] F. Jelinek, "Buffer overflow in variable length coding of fixed rate sources," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 490–501, 1968.
- [18] G. Longo, "A noiseless coding theorem for sources having utilities," *SIAM Journal on Applied Mathematics*, vol. 30, no. 4, pp. 739–748, 1976.
- [19] C. Arndt, *Information Measure-Information and Its Description in Science and Engineering*, Springer, Berlin, Germany, 2001.

