Hindawi
International Journal of Mathematics and Mathematical Sciences
Volume 2017, Article ID 2683293, 5 pages
https://doi.org/10.1155/2017/2683293

Research Article

A Joint Representation of Rényi's and Tsalli's Entropy with Application in Coding Theory

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 22 July 2017; Accepted 10 September 2017; Published 15 October 2017

Academic Editor: Hans Engler

We introduce a quantity which is called Rényi's-Tsalli's entropy of order ξ and discussed some of its major properties with Shannon and other entropies in the literature. Further, we give its application in coding theory and a coding theorem analogous to the ordinary coding theorem for a noiseless channel is proved. The theorem states that the proposed entropy is the lower bound of mean code word length.

1. Introduction

Let $\Delta_n = \{A = (a_1, a_2, \ldots, a_n) : a_i \geq 0,\ i = 1, 2, \ldots, n;\ \sum_{i=1}^{n} a_i = 1\}$, $n \geq 2$, be set of n-complete probability distributions. For any probability distribution $A = (a_1, a_2, \ldots, a_n) \in \Delta_n$, Shannon [1] defined an entropy given as

$$H(A) = -\sum_{i=1}^{n} (a_i) \log (a_i), \tag{1}$$

where the convention $0 \log(0) = 0$ is adopted (see Shannon [1]). Throughout this paper, logarithms are taken to the base $D (D > 1)$. A number of parametric generalizations of Shannon entropy are proposed by many authors in literature which produces (1) for specific values of parameters. The presence of parameters makes an entropy more flexible from application point of view. One of the first generalizations of (1) was proposed by Rényi [2] as

$$H_{Re}(A; \xi) = \frac{1}{1-\xi} \log \left(\sum_{i=1}^{n} a_i^{\xi} \right); \quad \xi > 0 \, (\neq 1). \tag{2}$$

Another well-known entropy was proposed by Havrda and Charvát [3]

$$H_{hc}(A; \xi) = \left(2^{1-\xi} - 1\right)^{-1} \left(\sum_{i=1}^{n} a_i^{\xi} - 1 \right); \quad \xi > 0 \, (\neq 1). \tag{3}$$

Independently, Tsalli [4] proposed another parametric generalization of the Shannon entropy as

$$H_T(A; \xi) = \frac{1}{1-\xi} \left(\sum_{i=1}^{n} a_i^{\xi} - 1 \right); \quad \xi > 0 \, (\neq 1). \tag{4}$$

Equations (3) and (4) essentially have the same expression except the normalized factor. The Havrda and Charvát entropy is normalized to 1. That is, if $A = (1/2, 1/2)$ then $H_{hc}(A; \xi) = 1$ whereas Tsalli's entropy is not normalized. Both the entropies yield the same result and we call these entropies as Tsalli-Havrda-Charvát entropy. Equations (2), (3), and (4) reduce to (1) when $\xi \to 1$.

N. R. Pal and S. K. Pal [5, 6] have proposed an exponential entropy as

$$H_{pp}(A) = \sum_{i=1}^{n} a_i \left(e^{1-a_i} - 1 \right). \tag{5}$$

These authors claim that the exponential entropy has some advantage over Shannon's entropy, especially within context of image processing. One such claim is that the exponential entropy has a fixed upper bound such as that for uniform distribution $(1/n, 1/n, \ldots, 1/n)$ and for the entropy in (5).

$$\lim_{n \to \infty} H_{pp}\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n} \right) = e - 1, \tag{6}$$

as compared to infinite limit (as $n \to \infty$) for the entropies in (1) and (2) and also for that in (3) when $\xi \in (0,1)$. Equation (5) was further generalized by Kvalseth [7] introducing a parameter as

$$H_K(A;\xi) = \frac{1}{\xi}\sum_{i=1}^{n} a_i\left(e^{1-a_i^{\xi}} - 1\right); \quad \xi \in \mathbb{R}. \tag{7}$$

In this paper, we introduce and study a new information measure which is called Rényi's-Tsalli's entropy of order ξ and a new mean code word length and discuss the relation with each other. In Section 2, Rényi's and Tsalli's entropy is introduced and also some of its major properties are discussed. In Section 3, the application of proposed information measure in coding theory is given and it is proved that the proposed information measure is the lower bound of mean code word length.

Now, in the next section, we propose a new parametric information measure.

2. A New Generalized Information Measure

However, in literature of information theory, there exists various generalizations of Shannon entropy; we introduce a new information measure as

$$H^{\xi}(A) = \begin{cases} \dfrac{1}{\xi^{-1}-\xi}\left[\log\left(\displaystyle\sum_{i=1}^{n} a_i^{\xi}\right) + \displaystyle\sum_{i=1}^{n} a_i^{\xi} - 1\right], & \xi > 0\ (\neq 1); \\ -\displaystyle\sum_{i=1}^{n} (a_i)\log(a_i), & \xi = 1. \end{cases} \tag{8}$$

Second case in (8) is a well-known Shannon entropy.

The quantity (8) introduced in the present section is a joint representation of Rényi's and Tsalli's entropy of order ξ. Such a name will be justified, if it shares some major properties with Shannon entropy and other entropies in the literature. We study some such properties in the next theorem.

2.1. Properties of Proposed Entropy

Theorem 1. *The parametric entropy $H^{\xi}(A)$, $\{A = (a_1, a_2, \ldots, a_n),\ 0 < a_i \le 1, \sum_{i=1}^{n} a_i = 1\}$ has the following properties.*

(1) Symmetry. $H^{\xi}(a_1, a_2, \ldots, a_n)$ is a symmetric function of $(a_1, a_2, \ldots, a_n)$.

(2) Nonnegative. $H^{\xi}(A) \ge 0$ for all $\xi > 0 (\neq 1)$.

(3) Expansible

$$H^{\xi}(a_1, a_2, \ldots, a_n; 0) = H^{\xi}(a_1, a_2, \ldots, a_n). \tag{9}$$

(4) Decisive

$$H^{\xi}(0,1) = H^{\xi}(1,0) = 0. \tag{10}$$

(5) Maximality

$$\begin{aligned} H^{\xi}(a_1, a_2, \ldots, a_n) &\le H^{\xi}\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \\ &= \frac{1}{\xi^{-1}-\xi}\left[\log\left(n^{1-\xi}\right) + n^{1-\xi} - 1\right]. \end{aligned} \tag{11}$$

(6) Concavity. The entropy $H^{\xi}(A)$ is a concave function for $0 < \xi < 1$ of the probability distribution $A = (a_1, a_2, \ldots, a_n)$, $a_i \ge 0$; $\sum_{i=1}^{n} a_i = 1$.

(7) Continuity. $H^{\xi}(a_1, a_2, \ldots, a_n)$ is continuous in the region $a_i \ge 0$ for all $i = 1, 2, \ldots, n$ and $\xi > 0 (\neq 1)$.

Proof. The properties (1), (3), (4), and (5) follow immediately from the definition. For property (7), we know that

$$\log\left(\sum_{i=1}^{n} a_i^{\xi}\right) + \sum_{i=1}^{n} a_i^{\xi} - 1 \tag{12}$$

is continuous in the region $a_i \ge 0$ for all $\xi > 0$. Thus, $H^{\xi}(A)$, is also continuous in the region $a_i \ge 0$ and $\xi > 0 (\neq 1)$ and $i = 1, 2, \ldots, n$.

Property (2)

Case 1 $(0 < \xi < 1)$

$$\sum_{i=1}^{n} a_i^{\xi} \ge 1. \tag{13}$$

From (13), we get

$$\log\left(\sum_{i=1}^{n} a_i^{\xi}\right) + \sum_{i=1}^{n} a_i^{\xi} - 1 \ge 0, \tag{14}$$

since $0 < \xi < 1 \Rightarrow \xi^{-1} - \xi > 0$.

Therefore, we get

$$\frac{1}{\xi^{-1}-\xi}\left[\log\left(\sum_{i=1}^{n} a_i^{\xi}\right) + \sum_{i=1}^{n} a_i^{\xi} - 1\right] \ge 0; \tag{15}$$

that is, $H^{\xi}(A) \ge 0$.

Therefore, we conclude that $H^{\xi}(A) \ge 0$ for all $0 < \xi < 1$.

Case 2 $(1 < \xi < \infty)$. The proof is on the same lines as in Case 1. (Note that inequality in (14) will get reversed for $1 < \xi < \infty$.)

Property (6). Now, we prove that $H^{\xi}(A)$ is a concave function of $A \in \Delta_n$.

Differentiating (8) twice with respect to a_i, we get

$$\frac{\partial^2 H^{\xi}(A)}{\partial a_i^2} = \frac{\xi}{\xi^{-1}-\xi}\left(\frac{(\xi-1)\sum_{i=1}^{n}\left(a_i^{\xi}\right)\sum_{i=1}^{n}\left(a_i^{\xi-2}\right)\left(\sum_{i=1}^{n}a_i^{\xi}+1\right)-\xi\left(\sum_{i=1}^{n}a_i^{\xi-1}\right)^2}{\left(\sum_{i=1}^{n}a_i^{\xi}\right)^2}\right). \tag{16}$$

Now, for $0<\xi<1$,

$$(\xi-1)\sum_{i=1}^{n}\left(a_i^{\xi}\right)\sum_{i=1}^{n}\left(a_i^{\xi-2}\right)\left(\sum_{i=1}^{n}a_i^{\xi}+1\right)-\xi\left(\sum_{i=1}^{n}a_i^{\xi-1}\right)^2 < 0, \quad \frac{\xi}{\xi^{-1}-\xi} > 0. \tag{17}$$

This implies that $H^{\xi}(A)$ is a concave function of $A\in\Delta_n$. □

3. A Measure of Length

Let a finite set of n input symbols $X=(x_1,x_2,\dots,x_n)$ be encoded using alphabet of D symbols; then it has been shown by Feinstein [8] that there is a uniquely decipherable code with lengths $N_1,N_2,\dots,N_n$ if and only if Kraft's inequality holds; that is,

$$\sum_{i=1}^{n}D^{-N_i}\le 1, \tag{18}$$

where D is the size of code alphabet. Furthermore, if

$$L=\sum_{i=1}^{n}N_i a_i \tag{19}$$

is the average code word length, then for a code satisfying (18), the inequality

$$L\ge H(A) \tag{20}$$

is also fulfilled and the equality $L=H(A)$ holds if and only if

$$N_i=-\log_D(a_i);\quad \forall i=1,2,\dots,n, \qquad \sum_{i=1}^{n}D^{-N_i}=1. \tag{21}$$

If $L<H(A)$, then by being suitably encoded into words of long sequences, the average length can be made arbitrarily close to $H(A)$ (see Feinstein [8]). This is Shannon's noiseless coding theorem. By considering Rényi's entropy [2], a coding theorem analogous to the above noiseless coding theorem has been established by Campbell [9] and the authors obtained bounds for it in terms of $H_{Re}(A;\xi)$.

Kieffer [10] defined class rules and showed $H_{Re}(A;\xi)$ is the best decision rule for deciding which of the two sources can be coded with expected cost of sequence of length N when $N\to\infty$, where the cost of encoding a sequence is assumed to be a function of length only. Further, in Jelinek [11], 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.

There are many different codes whose lengths satisfy the constraints (18). To compare different codes and pick out an optimum code it is customary to examine the mean length, $\sum_{i=1}^{n}N_i a_i$, and to minimize this quantity. This is a good procedure if the cost of using a sequence of length N_i is directly proportional to N_i. However, there may be occasions when the cost is more nearly an exponential function of N_i. This could be the case, for example, if the cost of encoding and decoding equipment was an important factor. Thus, in some circumstances, it might be more appropriate to choose a code which minimizes the quantity

$$C=\left[\xi\log_D\left(\sum_{i=1}^{n}a_i D^{-N_i((\xi-1)/\xi)}\right)+\left(\sum_{i=1}^{n}a_i D^{-N_i(\xi-1)/\xi}\right)^{\xi}-1\right];\quad \xi>0\,(\neq 1), \tag{22}$$

where ξ is a parameter related to the cost. For reasons which will become evident later we prefer to minimize a monotonic function of C. Clearly, this will minimize C.

In order to make the result of this paper more directly comparable with the usual coding theorem we introduce a quantity which resembles the mean length. Let a code length of order ξ be defined by

$$L^{\xi}(A)=\frac{1}{\xi^{-1}-\xi}\left[\xi\log_D\left(\sum_{i=1}^{n}a_i D^{-N_i((\xi-1)/\xi)}\right)+\left(\sum_{i=1}^{n}a_i D^{-N_i(\xi-1)/\xi}\right)^{\xi}-1\right],\quad \xi>0\,(\neq 1). \tag{23}$$

Remark 2. If $\xi=1$, then (23) becomes the well-known result studied by Shannon.

Remark 3. If all N_i are the same, say, $N_i=N$, then (23) becomes

$$L^{\xi}(P)=\frac{1}{\xi^{-1}-\xi}\left[N(1-\xi)+1-D^{N(1-\xi)}\right];\quad \xi>0\,(\neq 1). \tag{24}$$

This is reasonable property for any measure of length to possess.

Table 1: $\xi > 0 (\neq 1)$ (taking $\xi = 0.5$). Relation between the entropy $H^{\xi}(A)$ and the average code word length $L^{\xi}(A)$. N_i denote the lengths of Huffman code words, $\eta = H^{\xi}(A)/L^{\xi}(A)$ is the efficiency code, and $\nu = 1 - \eta$ is the redundancy of the code.

Relation between $H^{\xi}(A)$ and $L^{\xi}(A)$							
Length of Huffman code words (N_i)	Huffman code words	a_i	ξ	$H^{\xi}(A)$	$L^{\xi}(A)$	Efficiency code $\eta = \frac{H^{\xi}(A)}{L^{\xi}(A)}$	Redundancy code $\nu = 1 - \eta$
2	00	0.3					
2	10	.25					
2	11	0.2	0.5	1.7228	1.8707	0.9209	0.0791
3	011	0.1					
4	0100	0.1					
4	0101	.05					

In the following theorem, we give a lower bound for $L^{\xi}(A)$ in terms of $H^{\xi}(A)$.

Theorem 4. *If $N_1, N_2, \ldots, N_n$, denote the length of a uniquely decipherable code satisfying (18); then*

$$L^{\xi}(A) \geq H^{\xi}(A). \tag{25}$$

Proof. By Hölder's inequality,

$$\left[\sum_{i=1}^{n} x_i^p\right]^{1/p} \left[\sum_{i=1}^{n} y_i^q\right]^{1/q} \leq \sum_{i=1}^{n} x_i y_i, \tag{26}$$

for all $x_i, y_i > 0$, $i = 1, 2, \ldots, n$, and $1/p + 1/q = 1$; $p < 1(\neq 0)$; $q < 0$; or $q < 1(\neq 0)$; $p < 0$; equality holds if and only if, for some c, $x_i^p = c y_i^q$. Note that the direction of Hölder's inequality is reverse of the usual one for $p < 1$. (Beckenbach and Bellman [12], see p. 19). Making the substitutions, $p = (\xi - 1)/\xi$; $q = 1 - \xi$; $x_i = a_i^{\xi/(\xi-1)} D^{-N_i}$; $y_i = a_i^{\xi/(1-\xi)}$, in (26) and simplifying using (18). The following cases arise.

Case 1 (*when* $\xi > 1$)

$$\left[\sum_{i=1}^{n} a_i D^{-N_i((\xi-1)/\xi)}\right] \leq \left[\sum_{i=1}^{n} a_i^{\xi}\right]^{1/\xi}. \tag{27}$$

From (27), we get

$$\begin{aligned} &\xi \log_D \left(\sum_{i=1}^{n} a_i D^{-N_i((\xi-1)/\xi)}\right) + \left(\sum_{i=1}^{n} a_i D^{-N_i((\xi-1)/\xi)}\right)^{\xi} \\ &\quad - 1 \leq \log_D \left(\sum_{i=1}^{n} a_i^{\xi}\right) + \left(\sum_{i=1}^{n} a_i^{\xi}\right) - 1. \end{aligned} \tag{28}$$

Also,

$$\xi^{-1} - \xi < 0 \quad \text{for } \xi > 1. \tag{29}$$

Thus, from (28) and (29), we may conclude that $H^{\xi}(A) \leq L^{\xi}(A)$.

Case 2 (*when* $0 < \xi < 1$)

$$\left[\sum_{i=1}^{n} a_i D^{-N_i((\xi-1)/\xi)}\right] \geq \left[\sum_{i=1}^{n} a_i^{\xi}\right]^{1/\xi}. \tag{30}$$

From (30), we get

$$\begin{aligned} &\xi \log_D \left(\sum_{i=1}^{n} a_i D^{-N_i((\xi-1)/\xi)}\right) + \left(\sum_{i=1}^{n} a_i D^{-N_i((\xi-1)/\xi)}\right)^{\xi} \\ &\quad - 1 \geq \log_D \left(\sum_{i=1}^{n} a_i^{\xi}\right) + \left(\sum_{i=1}^{n} a_i^{\xi}\right) - 1. \end{aligned} \tag{31}$$

Also,

$$\xi^{-1} - \xi > 0 \quad \text{for } 0 < \xi < 1. \tag{32}$$

From (31) and (32), we get

$$H^{\xi}(A) \leq L^{\xi}(A). \tag{33}$$

Case 3. It is clear that the equality in (25) is valid if $N_i = -\log_D(a_i^{\xi} / \sum_{i=1}^{n} a_i^{\xi})$. The necessity of this condition for equality in (25) follows from the condition for equality in Hölder's inequality: in the case of reverse Hölder's equality given above, equality holds if and only if for some c,

$$x_i^p = c y_i^q, \quad i = 1, 2, \ldots, n. \tag{34}$$

Plugging this condition into our situation, with the x_i, y_i and p, q as specified, and using the fact that $\sum_{i=1}^{n} a_i = 1$, the necessity is true one. This proves the theorem. □

Remark 5. Huffman [13] introduced a measure for designing a variable length source code which achieves performance close to Shannon's entropy bound. For individual code word lengths N_i, the average length $\overline{L} = \sum_{i=1}^{n} a_i N_i$ of Huffman code is always within one unit of Shannon's measure of entropy; that is, $H(A) \leq \overline{L} < H(A)+1$, where $H(A) = -\sum_{i=1}^{n} a_i \log_2(a_i)$ is the Shannon's measure of entropy. Huffman coding scheme can also be applied to code word length $L^{\xi}(A)$ for code word length N_i; the average length $L^{\xi}(A)$ of Huffman code satisfies

$$L^{\xi}(A) \geq H^{\xi}(A). \tag{35}$$

In Table 1, we have developed the relation between the entropy $H^{\xi}(A)$ and average code word length $L^{\xi}(A)$.

Table 2: Computed values $L^{\xi}(A)$ with respect to ξ ($\xi < 1$).

ξ	.1	.2	.3	.4	.5	.6	.7	.8	.9
$L^{\xi}(A)$	1.25	1.67	1.78	1.82	1.87	1.92	1.96	1.99	2.01

Table 3: Computed values $L^{\xi}(A)$ with respect to ξ ($\xi > 1$).

ξ	10	20	30	40	50	60	70	80	90	100
$L^{\xi}(A)$	2.1618	2.2037	2.2199	2.2285	2.2337	2.2373	2.2399	2.2418	2.2433	2.2446

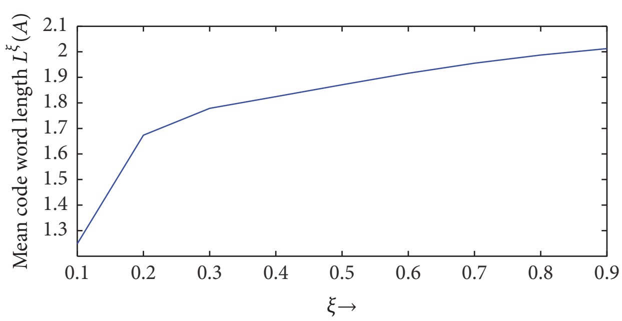

Figure 1: Monotonic behaviour of mean code word length $L^{\xi}(A)$ ($\xi < 1$).

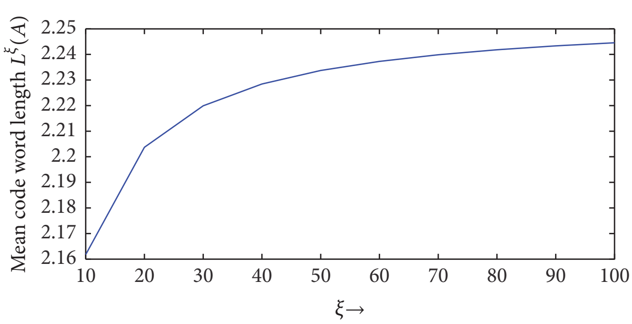

Figure 2: Monotonic behaviour of mean code word length $L^{\xi}(A)$ ($\xi > 1$).

From the Table 1, we can observe that average code word length $L^{\xi}(A)$ exceeds the entropy $H^{\xi}(A)$.

4. Monotonic Behaviour of Mean Code Word Length

In this section, we study the monotonic behaviour of mean code word length (23) with respect to parameter ξ. Let $P = (0.3, 0.25, 0.2, 0.1, 0.1, 0.05)$ be the set of probabilities. For different values of ξ, the calculated values of $L^{\xi}(A)$ are displayed in Tables 2 and 3.

Graphical representation of monotonic behaviour of $L^{\xi}(A)$ for ($\xi < 1$) is shown in Figure 1.

Graphical representation of monotonic behaviour of $L^{\xi}(A)$ for ($\xi > 1$) is shown in Figure 2.

Figures 1 and 2 explain the monotonic behaviour of $L^{\xi}(A)$ for $\xi < 1$ and $\xi > 1$, respectively. From the figures, it is clear that $L^{\xi}(A)$ is monotonically increasing for $\xi < 1$ as well as $\xi > 1$.

Conflicts of Interest

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

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.

[2] 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.

[3] J. Havrda and F. S. Charvát, "Quantification method of classification processes. Concept of structural α-entropy," *Kybernetika*, vol. 3, pp. 30–35, 1967.

[4] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.

[5] N. R. Pal and S. K. Pal, "Object-background segmentation using new definitions of entropy," *IEE Proceedings Part E Computers and Digital Techniques*, vol. 136, no. 4, pp. 284–295, 1989.

[6] N. R. Pal and S. K. Pal, "Entropy: a new definition and its applications," *The Institute of Electrical and Electronics Engineers Systems, Man, and Cybernetics Society*, vol. 21, no. 5, pp. 1260–1270, 1991.

[7] T. O. Kvalseth, "On exponential entropies," in *Proceedings of the IEEE International Conference on Systems, Man And Cybernatics*, vol. 4, pp. 2822–2826, 2000.

[8] A. Feinstein, *Foundations of Information Theory*, McGraw-Hill, New York, NY, USA, 1956.

[9] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965.

[10] 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.

[11] 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.

[12] E. F. Beckenbach and R. Bellman, *Inequalities*, Springer, New York, NY, USA, 1961.

[13] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.