Journal of Mathematical Extension
Vol. XX, No. XX, (2014), pp-pp (Will be inserted by layout editor)
ISSN: 1735-8299
URL: http://www.ijmex.com

An improvement of the upper bound on the entropy of information sources

Yamin Sayyari*
Department of Mathematics, Sirjan University Of Technology, Sirjan, Iran

Abstract. Theory of zeta functions and fractional calculus plays an important role in the statistical problems and Shannon's entropy. Estimation of Shannon's entropies of information sources from numerical simulation of long orbits is difficult. Our aim within this paper is to present a strong upper bound for the Shannon's entropy of information sources.

AMS Subject Classification: 37B40; 94A17; 26A51.
Keywords and Phrases: Entropy, Shannon's entropy, Information source, Stochastic process, zeta function.

1 Introduction

If $s > 1$, then Riemann function is defined as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

The subject of fractional calculus has emerged as a powerful mathematical instrument during the past years, and is used in every branch of

Received: XXXX; Accepted: XXXX (Will be inserted by editor)
*Corresponding Author

the statistics, engineering, and in other fields. S. Golomb showed that Riemann's zeta function ζ induces a probability distribution $\pi(n) = \frac{n^{-s}}{\zeta(s)}$ on the positive integers, for every $s > 1$ [7]. In Guiasu [9], the author proved that the probability distribution mentioned above is the unique solution of an entropy-maximization problem. Fractional calculus of zeta functions can also be used to maximize

$$H = -\sum_n \pi(n) \log \pi(n),$$

where $\{\pi(n) : n \in \mathbb{N}\}$ is a probability distribution on $\mathbb{N}$ [8].

Theorem 1.1. *[8] Let $\alpha \in \mathbb{R} \backslash \mathbb{Z}$, $\pi(n) > 0$ and $\sum_n \pi(n) = 1$. The maximization of Shannon entropy $H = -\sum_n \pi(n) \log \pi(n)$ and*

$$\sum_{n \in \mathbb{N}} \pi(n) \log D_f^\alpha n^{-x} = \chi_\alpha, \quad x > 1 + \alpha,$$

has a solution given by

$$\pi(n) = \frac{D_f^\alpha n^{-x}}{\zeta^{(\alpha)(x)}}, \quad n \in \mathbb{N}.$$

where the forward Grunwald-Letnikov fractional derivative of f is defined as follows:

$$D_f^\alpha f(x) = \lim_{h \to 0^+} \frac{\sum_{m=0}^{\infty} \binom{\alpha}{m} (-1)^m f(x - mh)}{h^\alpha}.$$

Entropy and mutual information for random variables play important roles in dynamical systems and information theory. The entropy actually measures the degree of irregularities of a dynamic system, and researchers have done so much to calculate this concept, which is often successful [4, 12], but numerical calculations of entropy are still difficult. Tapus and Popescu presented a strong upper bound for the classical Shannon entropy [11]. In [11, 16, 14], the authors presented a strong upper bound for the classical Shannon entropy. In [13], the authors presented the algebraic and Shannon entropies for hypergroupoids and commutative hypergroups, respectively, and studies their fundamental properties. In [14], the author applying Jensen's inequality in information theory and we obtain some results for the Shannon's entropy of random variables and Shannon's entropy of information sources. Our purpose within this work is to present a strong upper bound for the Shannon entropy of information sources, refining recent results from the literature.

Let X be a non-empty set, $\mathcal{F}$ is an σ-algebra of subsets of X, μ is a measure on X and $\mu(X) = 1$, then $(X, \mathcal{F}, \mu)$ is called measure probability space. A finite set of measurable sets $\alpha = \{A_1, \ldots, A_n\}$ is called a finite partition if the following properties are fulfilled [17]:

$$\bigcup_{i=1}^{n} A_i = X, \quad \text{and} \quad A_i \cap A_j = \emptyset \quad \text{for every } i, j (1 \leq i \neq j \leq n).$$

For a partition $\alpha = \{A_1, \ldots, A_n\}$, the entropy of α is defined by

$$H_\mu(\alpha) := -\sum_{i=1}^{n} \mu(A_i) \log(\mu(A_i)).$$

Definition 1.2. [5] Let S be a random variable on X with discrete finite state space $A = \{a_1, ..., a_N\}$. We define $p : A \to [0, 1]$ by $p(s) = \mu\{\omega \in X : S(\omega) = s\}$. The Shannon's entropy of S is defined by

$$H_\mu(S) := -\sum_{s \in A,\ p(s) \neq 0} p(s) \log p(s).$$

An information sources $\mathbf{S}$ is a sequence $(S_n)_{n=1}^{\infty}$ of the random variables $S_n : X \longrightarrow A$, where $n \in \mathbb{N}$. For given $L \geq 1$ we define a mapping $p : A^L \to [0, 1]$ by $p(s_1^L) = \mu\{\omega \in X : S_1(\omega) = s_1, ..., S_L(\omega) = s_L\}$. The Shannon entropy of order L and the Shannon entropy of source $\mathbf{S}$ are respectively defined by

$$H_\mu(S_1^L) = -\frac{1}{L} \sum_{s_1^L \in A^L} p(s_1, ..., s_L) \log p(s_1, ..., s_L), \text{ and } h_\mu(\mathbf{S}) = \lim_{L \to \infty} H_\mu(S_1^L).$$

where the the summation is taken over the collection $\{s_1^L \in A^L : p(s_1^L) \neq 0\}$. In this paper we use the symbol s_1^L instead of notation $(s_1, ..., s_L)$ and Let $p(s_1^L) \neq 0$ for every $L \in \mathbb{N}$.

Theorem 1.3. *[14] Let $I = [a, b]$ be an interval, $H : A^L \longrightarrow I$ be a function, and $f : I \longrightarrow \mathbb{R}$ be a convex function, then*

$$\begin{aligned}
&\sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L)) \\
&\geq \max\{p(r_1^L) f(H(r_1^L)) + p(t_1^L) f(H(t_1^L)) \\
&- (p(r_1^L) + p(t_1^L)) f(\frac{p(r_1^L) H(r_1^L) + p(t_1^L) H(t_1^L)}{p(r_1^L) + p(t_1^L)})\}, \qquad (1)
\end{aligned}$$

where the maximum is taken over all $r_1^L \neq t_1^L \in A^L$.

2 main results

In this section, we continue with a refinement of Theorem 1.3, as follows:

Theorem 2.1. *Let $I = [a, b]$ be an interval, $H : A^L \longrightarrow I$ be a function, and $f : I \longrightarrow \mathbb{R}$ be a convex function, then*

$$\begin{aligned}
&\sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L)) \\
&\geq \max\{p(r_1^L) f(H(r_1^L)) + p(t_1^L) f(H(t_1^L)) + p(u_1^L) f(H(u_1^L)) \\
&- (p(r_1^L) + p(t_1^L) + p(u_1^L)) f(\frac{p(r_1^L) H(r_1^L) + p(t_1^L) H(t_1^L) + p(u_1^L) H(u_1^L)}{p(r_1^L) + p(t_1^L) + p(u_1^L)})\}, \\
&\geq \max\{p(r_1^L) f(H(r_1^L)) + p(t_1^L) f(H(t_1^L)) + p(u_1^L) f(H(u_1^L))\} \\
&- (p(r_1^L) + p(t_1^L) + p(u_1^L)) f(\frac{p(r_1^L) H(r_1^L) + p(t_1^L) H(t_1^L) + p(u_1^L) H(u_1^L)}{p(r_1^L) + p(t_1^L) + p(u_1^L)})\},
\end{aligned}$$

where the maximum is taken over all distinct $r_1^L, t_1^L, u_1^L \in A^L$.

Proof. Choose arbitrary $t_1^L, r_1^L, u_1^L \in A^L$. So,

$$f(\sum_{s_1^L \in A^L} p(s_1^L)H(s_1^L)) = f(\sum_{s_1^L \neq r_1^L, t_1^L, u_1^L \in A^L} p(s_1^L)H(s_1^L))$$
$$+ (p(r_1^L) + p(t_1^L) + p(u_1^L))(\frac{p(r_1^L)H(r_1^L) + p(t_1^L)H(t_1^L) + p(u_1^L)H(u_1^L)}{p(r_1^L) + p(t_1^L) + p(u_1^L)})$$
$$\leq \sum p(s_1^L)f(H(s_1^L))$$
$$+ (p(r_1^L) + p(t_1^L) + p(u_1^L))f(\frac{p(r_1^L)H(r_1^L) + p(t_1^L)H(t_1^L) + p(u_1^L)H(u_1^L)}{p(r_1^L) + p(t_1^L) + p(u_1^L)}),$$

where $s_1^L \neq r_1^L, t_1^L, u_1^L \in A^L$. Therefore,

$$\sum_{s_1^L \in A^L} p(s_1^L)f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L)H(s_1^L))$$
$$\geq p(r_1^L)f(H(r_1^L)) + p(t_1^L)f(H(t_1^L)) + p(u_1^L)f(H(u_1^L))$$
$$- (p(r_1^L) + p(t_1^L) + p(u_1^L))f(\frac{p(r_1^L)H(r_1^L) + p(t_1^L)H(t_1^L) + p(u_1^L)H(u_1^L)}{p(r_1^L) + p(t_1^L) + p(u_1^L)}).$$

Since $s_1^L, t_1^L \in A^L, u_L^1$ are arbitrary,

$$\sum_{s_1^L \in A^L} p(s_1^L)f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L)H(s_1^L))$$
$$\geq \max\{p(r_1^L)f(H(r_1^L)) + p(t_1^L)f(H(t_1^L)) + p(u_1^L)f(H(u_1^L))\}$$
$$- (p(r_1^L) + p(t_1^L) + p(u_1^L))f(\frac{p(r_1^L)H(r_1^L) + p(t_1^L)H(t_1^L) + p(u_1^L)H(u_1^L)}{p(r_1^L) + p(t_1^L) + p(u_1^L)})\},$$

where the maximum is taken over all distinct $r_1^L, t_1^L, u_1^L \in A^L$. On the other hand,

$$f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)+p(u_1^L)H(u_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)})$$
$$=f(\frac{p(r_1^L)+p(t_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)}\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)}+\frac{p(u_1^L)H(u_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)})$$
$$\leq \frac{p(r_1^L)+p(t_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)}f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)})$$
$$+\frac{p(u_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)}f(H(u_1^L)).$$

So,

$$(p(r_1^L)+p(t_1^L)+p(u_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)+p(u_1^L)H(u_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)})$$
$$\leq (p(r_1^L)+p(t_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)})+(p(u_1^L))f(H(u_1^L)).$$

Thus,

$$p(r_1^L)f(H(r_1^L))+p(t_1^L)f(H(t_1^L))+p(u_1^L)f(H(u_1^L))$$
$$-(p(r_1^L)+p(t_1^L)+p(u_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)+p(u_1^L)H(u_1^L)}{p(r_1^L)+p(t_1^L)+p(u_1^L)})$$
$$\geq p(r_1^L)f(H(r_1^L))+p(t_1^L)f(H(t_1^L))\}$$
$$-(p(r_1^L)+p(t_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)}),$$

which completes the proof. □

In order to present the generalization, we define some notation, as follows:

$$T_k := \max\{\sum_{i=1}^{k} p({r_i}_1^L)f(H({r_i}_1^L)) - (\sum_{i=1}^{k} p({r_i}_1^L))f(\frac{\sum_{i=1}^{k} p({r_i}_1^L)H({r_i}_1^L)}{\sum_{i=1}^{k} p({r_i}_1^L)})\}$$

where $2 \leq k \leq N^L - 1$, the maximum is taken over all distinct ${r_i}_1^L \in A^L$.

Theorem 2.2. *Let $I = [a, b]$ be an interval, $H : A^L \longrightarrow I$ be a function, $|A| = N$ and $f : I \longrightarrow \mathbb{R}$ be a convex function, then*

$$0 \leq T_2 \leq T_3 \leq ... \leq T_{N^L-1} \leq \sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L)).$$

Proof. The proof is similar to the proof of Theorem 2.1. □

3 The sources entropy upper bound

In this section we present a strong upper bound for the Shannon's entropy of information sources.

Theorem 3.1. *Let* $\mathbf{S}$ *be an information source. Then*

$$h_\mu(\mathbf{S}) \leq \log N - \max_k \{ \lim_{L\to\infty} \frac{1}{L} log[\{ \frac{k}{\sum_{i=1}^k p(r_{i1}^L)} \}^{\sum_{i=1}^k p(r_{i1}^L)}] \times [\prod_{i=1}^k \{p(r_{i1}^L)\}^{p(r_{i1}^L)}]\}.$$

Proof. Since

$$- LH_\mu(S_1^L) + \log(N^L) \geq \max_k \{ -\sum_{i=1}^k p(r_{i1}^L) \log(\frac{1}{p(r_{i1}^L)}) + (\sum_{i=1}^k p(r_{i1}^L)) \times \log(\frac{k}{\sum_{i=1}^k p(r_{i1}^L)})\} = \max_k \{ \log(\prod_{i=1}^k \{p(r_{i1}^L)\}^{p(r_{i1}^L)}) + \log[\{ \frac{k}{\sum_{i=1}^k p(r_{i1}^L)} \}^{\sum_{i=1}^k p(r_{i1}^L)}]\},$$

$$\log N - H_\mu(S_1^L) \geq \max\{ \frac{1}{L} log[\{ \frac{k}{\sum_{i=1}^k p(r_{i1}^L)} \}^{\sum_{i=1}^k p(r_{i1}^L)}][\prod_{i=1}^k \{p(r_{i1}^L)\}^{p(r_{i1}^L)}]\},$$

and

$$H_\mu(S_1^L) \leq \log N - \max_k \{ \frac{1}{L} log[\{ \frac{k}{\sum_{i=1}^k p(r_{i1}^L)} \}^{\sum_{i=1}^k p(r_{i1}^L)}][\prod_{i=1}^k \{p(r_{i1}^L)\}^{p(r_{i1}^L)}]\}.$$

Therefore,

$$
\begin{aligned}
&h_\mu(\mathbf{S}) \\
&\leq \log N - \lim_{L\to\infty} \max_k \{ \frac{1}{L} log[\{\frac{k}{\sum_{i=1}^k p({r_i}_1^L)}\}^{\sum_{i=1}^k p({r_i}_1^L)}][\prod_{i=1}^k \{p({r_i}_1^L)\}^{p({r_i}_1^L)}]\} \\
&\leq \log N - \max_{2\leq k\leq N^L-1} \{ \lim_{L\to\infty} \frac{1}{L} log[\{\frac{k}{\sum_{i=1}^k p({r_i}_1^L)}\}^{\sum_{i=1}^k p({r_i}_1^L)}][\prod_{i=1}^k \{p({r_i}_1^L)\}^{p({r_i}_1^L)}]\}.
\end{aligned}
$$

□

Entropy of information sources is very important in synamical systems and information theory. Let $(X, \mathcal{F}, \mu)$ me a probability measure space. For a partition

$$\alpha = \{A_0, ..., A_N\}$$

and measure-preserving dynamical system $f : X \longrightarrow X$, the maps

$$S_n : X \longrightarrow T_N := \{0, ..., N\},$$

defined as

$$S_n(x) = i \text{ if and only if } f^n(x) \in A_i$$

are random variables on the probability measure space X. In this case we have

$$p(i) = \mu(A_i),$$

for every $i(0 \leq i \leq N)$, and $h_\mu(\mathbf{S}_\alpha) = h_\mu(f, \alpha)$ where $\mathbf{S}_\alpha = \{S_n\}$ [5]. Since The metric entropy of f is then the supremum of $h_\mu(f, \alpha)$ over all finite partitions of $(X, \mathcal{F}, \mu)$ (i.e.

$$h_\mu(f) = \sup_\alpha h_\mu(f, \alpha) = \sup_\alpha h_\mu(\mathbf{S}_\alpha).) \tag{2}$$

Thus, an approximation of entropy f is obtained by using 2.

4 motivation and conclusion

In this paper, we have obtained some mathematical inequalities for entropy of information sources. Also we found new and strong bounds for the Shannon's entropy of information sources. Theorem 3.1, shows that in general,

$$\log N - \frac{1}{L} log[\{\frac{k}{\sum_{i=1}^{k} p({r_i}_1^L)}\}^{\sum_{i=1}^{k} p({r_i}_1^L)}] \times [\prod_{i=1}^{k}\{p({r_i}_1^L)\}^{p({r_i}_1^L)}]$$

can only be expected to be an upper bound of $h_\mu(\mathbf{S})$, we will try to extend it in the future.

References

[1] M. Adeel, K.A. Khan, D. Pec̆arić, J. Pec̆arić, Generalization of the Levinson inequality with applications to information theorey, J. Inequal. Appl., 230 (2019).

[2] M. Adeel, K.A. Khan, D. Pec̆arić, J. Pec̆arić, Levinson type inequalities for higher order convex functions via Able-Gontscharo interpolation, Ad. Differ. Equ., 430 (2019).

[3] M. Adeel, K.A. Khan, D. Pec̆arić, J. Pec̆arić, Estimation of f-divergence and Shannon entropy by Levinson type inequalities via new Green's functions and Lidstone polynomial, Ad. Differ. Equ., 2020, 27 (2020).

[4] J.M. Amigo, M.B. Kennel, Topological permutation entropy. Physica D. 231.

[5] J.M. Amigo, Permutation Complexity in Dynamical Systems "Ordinal Patterns, Permutation Entropy, and All That", Springer-Verlag, Berlin, 2010.

[6] C. Cattani and E. Guariglia, Fractional derivative of the Hurwitz ζ-function and chaotic decay to zero, J. King Saud Univ. Sci. 28 (1) (2016), 75–81, https://doi.org/10.1016/j.jksus.2015.04.003.

[7] S.W. Golomb, A class of probability distributions on the integers, J. Number Theory, 2 (1970), 189-192.

[8] E. Guariglia, Fractional calculus, zeta functions and Shannon entropy, Open Mathematics, 19 (1) (2021), https://doi.org/10.1515/math-2021-0010.

[9] S. Guiasu, An optimization problem related to the zeta function, Canad. Math. Bull. 29 (1) (1986), 70–73, https://doi.org/10.4153/CMB-1986-013-7.

[10] R.L. Magin, C. Ingo, Entropy and information in a fractional order model of anomalous diffusion, 16th IFAC Symposium on System Identification The International Federation of Automatic Control Brussels, Belgium. (2012), 11-13.

[11] N. Tapus, P.G. Popescu, A new entropy upper bound, Appl. Math. Lett. 25 (11) (2012), 1887-1890.

[12] Ch. Corda, M. FatehiNia, M.R. Molaei, Y. Sayyari, Entropy of Iterated Function Systems and Their Relations with Black Holes and Bohr-Like Black Holes Entropies, Entropy. 20, 56, (2018).

[13] A. Mehrpooya, Y. Sayyari, M.R. Molaei, Algebraic and Shannon entropies of commutative hypergroups and their connection with information and permutation entropies and with calculation of entropy for chemical algebras, Soft Computing 23 (24) (2019), 13035-13053.

[14] Y. Sayyari, New bounds for entropy of information sources, Wavelets and Linear Algebra, 7 (2) (2020), 1-9.

[15] Y. Sayyari, New entropy bounds via uniformly convex functions, Chaos, Solitons and Fractals 141 (2020), 110360 (Doi.org/10.1016/j.chaos.2020.110360).

[16] S. Simic, Jensen's inequality and new entropy bounds, Appl. Math. Lett. 22 (8) (2009), 1262-1265.

[17] P. Walters, An introduction to ergodic theory (Springer Verlag, New York, 2000).

Yamin Sayyari
Department of Mathematics
Assistant Professor of Mathematics
Sirjan University Of Technology
Sirjan, Iran
E-mail: ysayyari@gmail.com