ON SHANNON ENTROPY BOUNDS

YAMIN SAYYARI

ABSTRACT. Entropy, has many applications in thermodynamics, code theory, physics, statistics and information theory. In this paper, we present some new and interesting results related to the bounds of the Shannon entropy.

Subject Classification: 37B40; 26B25; 94A17; 26D15; 26D20.
Keywords: Shannon entropy; Bounds, Refinements.

1. Introduction

Entropy plays an important role in many areas of mathematics, probability and physics. Shannon's entropy, as metric entropy, is in general difficult to calculate and even to estimate. See [1] for other methods to estimate the Shannon entropy and [**?**] for a review on entropy estimation. In [10, 12], the authors presented some bounds for the classical Shannon's entropy. The results of this paper improve the results in [4, 8, 9, 11, 12].

2. Basic notions

Let $p_1, ..., p_n$ be a positive weight sequence with $\sum_{i=1}^{n} p_i = 1$, and let $\overline{x} = \{x_1, ..., x_n\} \subseteq I := [a, b]$ be a sequence. The well-known Jensen's inequality states that: If f is convex on I, then $\sum_{i=1}^{n} p_i f(x_i) - f(\sum_{i=1}^{n} p_i x_i) \geq 0$. The sum $\sum_{i=1}^{n} p_i x_i$ is called the convex combination of x_i.

Lemma 2.1. [3] *Let f be a differentiable convex mapping. Then*

$$0 \leq \sum_{i=1}^{n} p_i f(x_i) - f(\sum_{i=1}^{n} p_i x_i) \leq \frac{1}{4}(b-a)(f'(b) - f'(a)) := D_f(a, b). \tag{2.1}$$

Dragomir's result (2.1), implies $0 \leq \log n - H(X) \leq \frac{(\nu-\mu)^2}{4\mu\nu} := D(\mu, \nu)$.

Proposition 2.2. [11] *For $\mu := \min_{1 \leq i \leq n}\{p_i\}$ and $\nu := \max_{1 \leq i \leq n}\{p_i\}$, have*

$$\begin{aligned} m(\mu, \nu) := \mu \log(\frac{2\mu}{\mu+\nu}) + \nu \log(\frac{2\nu}{\mu+\nu}) &\leq \log n - H(X) \\ &\leq \log(\frac{(\mu+\nu)^2}{4\mu\nu}) := M(\mu, \nu). \end{aligned} \tag{2.2}$$

Proposition 2.3. [11] *Under the notation of Proposition 2.2, have*

$$m(\mu, \nu) \leq \log n - H(X) \leq nm(\mu, \nu). \tag{2.3}$$

Date: Received: , Accepted: .

Proposition 2.4. [8] *Under the notation of Proposition 2.2, have*

$$\tilde{m}(\mu,\nu) \leq \log n - H(X) \leq \tilde{M}(\mu,\nu), \tag{2.4}$$

where $\tilde{m}(\mu,\nu) := m(\mu,\nu) + \frac{\mu^2(2-n\mu-n\nu)^2}{2(\mu+\nu)(1-\mu-\nu)}$, *and*

$$\tilde{M}(\mu,\nu) := M(\mu,\nu) - \frac{(\mu+\nu-2n\mu\nu)^2 + 2\mu\nu(1-\mu n)(\nu n-1)}{4\nu^2}.$$

Proposition 2.5. [8] *Let* $\mu := \min_{1\leq i\leq n}\{p_i\}$ *and* $\nu := \max_{1\leq i\leq n}\{p_i\}$. *Then*

$$\overline{m}(\mu,\nu) \leq \log n - H(X) \leq \overline{M}(\mu,\nu), \tag{2.5}$$

where $\overline{m}(\mu,\nu) := m(\mu,\nu) + \frac{(2-n\mu-n\nu)^2}{4\nu n(n-2)}$ *and*

$$\overline{M}(\mu,\nu) := nm(\mu,\nu) - \frac{(2-n\mu-n\nu)^2 + 2(n\nu-1)(1-n\mu)}{4\nu n}.$$

Theorem 2.6. [12] *If* $X = \{p_i\}_{i=1}^n$ *is a positive probability distribution, then*

$$H(X) \leq \log n - \max_{1\leq\mu_1<\ldots<\mu_{n-1}\leq n}\{log([(\frac{n-1}{\sum_{i=1}^{n-1} p_{\mu_k}})^{\sum_{k=1}^{n-1} p_{\mu_k}}][\prod_{k=1}^{n-1} p_{\mu_k}^{p_{\mu_k}}])\}.$$

Theorem 2.7. [4] *Let* $X = \{p_i\}_{i=1}^n$ *be a positive probability distribution and* $\mu = (\mu_1, .., \mu_n)$. *Then*

$$H(X) \leq \log(n) - \frac{1}{n}\sum_{i=1}^{n}(e^{1-np_i} - 1) - \max_{1\leq\mu_1<\ldots<\mu_{n-1}\leq n}\{F(\mu) + G(\mu)\},$$

where

$$F(\mu) = log([(\frac{n-1}{\sum_{i=1}^{n-1} p_{\mu_k}})^{\sum_{k=1}^{n-1} p_{\mu_k}}][\prod_{k=1}^{n-1} p_{\mu_k}^{p_{\mu_k}}]),$$

$$G(\mu) = \frac{n-1}{n}(e^{1-\frac{n}{n-1}\sum_{i=1}^{n-1} p_{\mu_i}} - 1) - \frac{1}{n}\sum_{i=1}^{n-1}(e^{1-np_{\mu_i}} - 1).$$

3. Main results

In this section we obtain new upper bounds for Shannon's entropy of a positive probability distribution.

Theorem 3.1. *Let* $X = \{p_1, ..., p_n\}$ *be a positive probability distribution and* $\mu_0 := \min_{1\leq i\leq n}\{p_i\}$, *then*

$$H(X) \leq$$

$$\log n - \max_{2\leq i\leq n-1}\{\max_{1\leq\mu_1<\ldots<\mu_i\leq n}\{F_i(\mu)\exp(\frac{\mu_0^2(n\sum_{k=1}^{i} p_{\mu_k} - i)^2}{2(1-\sum_{k=1}^{i} p_{\mu_k})(\sum_{k=1}^{i} p_{\mu_k})})\}\}.$$

where

$$F_i(\mu) := \log([(\frac{i}{\sum_{k=1}^{i} p_{\mu_k}})^{\sum_{k=1}^{i} p_{\mu_k}}][\prod_{k=1}^{i} p_{\mu_k}^{p_{\mu_k}}].$$

$2 \leq i \leq n-1$ *and* $\mu = (\mu_1, .., \mu_n)$.

Corollary 3.2. *Let* $X = \{p_1, ..., p_n\}$ *be a positive probability distribution and* $\mu_0 := \min_{1 \leq i \leq n}\{p_i\}$, *then*

$$H(X) \leq \log n - \max_{1 \leq \mu_1 < ... < \mu_{n-1} \leq n} \{F(\mu) \exp(\frac{\mu_0^2 (n \sum_{k=1}^{n-1} p_{\mu_k} - n + 1)^2}{2(1 - \sum_{k=1}^{n-1} p_{\mu_k})(\sum_{k=1}^{n-1} p_{\mu_k})}))\}.$$

where

$$F(\mu) := \log([(\frac{n-1}{\sum_{k=1}^{n-1} p_{\mu_k}})^{\sum_{k=1}^{n-1} p_{\mu_k}}][\prod_{k=1}^{n-1} p_{\mu_k}^{p_{\mu_k}}]$$

and $\mu = (\mu_1, .., \mu_n)$.

Note that, the estimation in Corollary 3.2 is better than the estimation in Theorem 2.6.

Theorem 3.3. *Let* $X = \{p_1, ..., p_n\}$ *be a positive probability distribution. Let* $\mu := \min_{1 \leq i \leq n}\{p_i\} = p_\alpha$, $\mu_1 := \min_i\{p_i : i \neq \alpha\}$, $\nu := \max_{1 \leq i \leq n}\{p_i\} = p_\beta$ *and* $\nu_1 := \max\{p_i : i \neq \beta\}$ *. Then*

$$0 \leq \log n - H(X) \leq \tilde{M}_1(\mu, \nu, \mu_1, \nu_1), \tag{3.1}$$

where $\tilde{M}_1(\mu, \nu, \mu_1, \nu_1) := \tilde{M}(\mu, \nu) - \frac{\mu^3(\nu_1 - \mu_1)^2}{(1 - \mu - \nu) 2 \nu_1^2 \mu_1}$.

Remark 3.4. Since $\tilde{M}_1(\mu, \nu, \mu_1, \nu_1) \leq \tilde{M}(\mu, \nu) \leq M(\mu, \nu) \leq D(\mu, \nu)$, the estimation (3.1) is better than (2.4) and (2.2).

Theorem 3.5. *Let* $X = \{p_1, ..., p_n\}$ *be a positive probability distribution. Let* $\mu := \min_{1 \leq i \leq n}\{p_i\} = p_\alpha$, $\mu_1 := \min_i\{p_i : i \neq \alpha\}$, $\nu := \max_{1 \leq i \leq n}\{p_i\} = p_\beta$ *and* $\nu_1 := \max\{p_i : i \neq \beta\}$ *. Then*

$$0 \leq \log n - H(X) \leq \overline{M}_1(\mu, \nu, \mu_1, \nu_1), \tag{3.2}$$

where $\overline{M}_1(\mu, \nu, \mu_1, \nu_1) := \overline{M}(\mu, \nu) - \frac{\mu \mu_1 (\nu_1 - \mu_1)^2}{2\nu(1 - \mu - \nu)}$.

Remark 3.6. Since $\overline{M}_1(\mu, \nu, \mu_1, \nu_1) \leq \overline{M}(\mu, \nu) \leq nm(\mu, \nu)$, the estimation (3.2) is better than (2.5) and (2.3).

Example 3.7. Let $n = 10^k, \mu = 10^{-k-1}$, $\nu = 10^{-k+1}(k > 2)$ and

$$X = \{10^{-k-1}, 10^{-k-1}, x_3, x_4, ..., x_{10^k - 2}, 10^{-k+1}, 10^{-k+1}\}.$$

Then $M(\mu,\nu) \simeq 1.406$, $\tilde{M}(\mu,\nu) \simeq 1.202058$. Since,

$$\begin{aligned}
\tilde{M}_1(\mu,\nu,\mu_1,\nu_1) &= \tilde{M}(\mu,\nu) - \frac{\mu^3(\nu_1-\mu_1)^2}{(1-\mu-\nu)2\nu_1^2\mu_1} \\
&= 1.202058 - \frac{10^{-3k-3}(10^{-k+1}-10^{-k-1})^2}{(1-10^{-k-1}-10^{-k+1})2\times 10^{-2k+2}10^{-k-1}} \\
&= 1.202058 - \frac{9.99^2}{2\times 10^4} \times \frac{10^{-2k}}{1-10^{-k+1}-10^{-k-1}} \\
&\leq 1.202058 - \frac{9.99^2}{2\times 10^4} \times 3 \times 10^{-2k} \\
&= 1.202058 - 0/0149 \times 10^{-2k},
\end{aligned}$$

$$0 \leq \log n - H(X) \leq 1.202058 - 14/9 \times 10^{-2k-3}.$$

Example 3.8. Let $n = 100^k, \mu = 100^{-k-1}$, $\nu = 100^{-k+1} (k > 2)$ and

$$X = \{100^{-k-1}, 100^{-k-1}, x_3, x_4, ..., x_{100^k-2}, 100^{-k+1}, 100^{-k+1}\}.$$

Then

$$nm(\mu,\nu) - \overline{M}(\mu,\nu) \simeq 24.5049.$$

Also,

$$\begin{aligned}
\overline{M}_1(\mu,\nu,\mu_1,\nu_1) &= \overline{M}(\mu,\nu) - \frac{\mu\mu_1(\nu_1-\mu_1)^2}{2\nu(1-\mu-\nu)} \\
&= \overline{M}(\mu,\nu) - \frac{100^{-k-1}\times 100^{-k-1}(100^{-k+1}-100^{-k-1})^2}{2100^{-k+1}(1-100^{-k-1}-100^{-k+1})} \\
&= \overline{M}(\mu,\nu) - \frac{99.99^2}{2\times 100^3} \times \frac{100^{-2k}}{1-100^{-k+1}-100^{-k-1}} \\
&\leq \overline{M}(\mu,\nu) - \frac{99.99^2}{2\times 100^3} \times 3 \times 100^{-2k} \\
&= \overline{M}(\mu,\nu) - 149/9 \times 100^{-2k-2}.
\end{aligned}$$

So, $0 \leq \log n - H(X) \leq \overline{M}(\mu,\nu) - 149/9 \times 100^{-2k-2}$.

References

[1] J.M. Amigo, Permutation complexity in dynamical systems (2010), Springer-Verlag, Berlin-Heidelberg.

[2] S.S. Dragomir (2010), A new refinement of Jensen's inequality in linear spaces with applications, Mathematical and Computer Modelling, 52 (10), 1497–1505.

[3] S.S. Dragomir (1999-2000), A converse result for Jensen's discrete inequality via Grüss inequality and applications in information theory, An. Univ. Oradea. Fasc. Mat. 7, 178-189.

[4] G. Lu (2018), A refined upper bound for entropy, University Politehnica Of Bucharest Scientific Bulletin-series A-applied Mathematics And Physics, 80 (2).

[5] G. Lu (2018), New refinements of Jensen's inequality and entropy upper bounds, Journal of Mathematical Inequalities, 12 (2), 403-421.

[6] Y. Sayyari, New refinements of Shannon's entropy upper bounds, Journal of Information and Optimization Sciences, (2021), 1-15.
[7] Y. Sayyari (2020), New bounds for entropy of information sources, Wavelet and Linear Algebra, 7 (2), 1-9.
[8] Y. Sayyari (2020), New entropy bounds via uniformly convex functions, Chaos, Solitons and Fractals, 141 (1), (DOI: 10.1016/j.chaos.2020.110360).
[9] Y. Sayyari, An improvement of the upper bound on the entropy of information sources, Journal of Mathematical Extension, Vol 15 (2021).
[10] S. Simic (2008), On a global bound for Jensen's inequality, J. Math. Anal. Appl., 343, 414-419 (DOI: 10.1016/j.jmaa.2008.01.060).
[11] S. Simic (2009), Jensen's inequality and new entropy bounds, Applied Mathematics Letters, 22 (8), 1262-1265.
[12] N. Tapus, P.G. Popescu (2012), A new entropy upper bound, Applied Mathematics Letters, 25 (11), 1887–1890.

(Yamin Sayyari) Department of Mathematics, Sirjan University Of Technology, Sirjan, Iran
Email address: `ysayyari@gmail.com, y.sayyari@sirjantech.ac.ir`