A REFINED UPPER BOUND FOR ENTROPY

Guoxiang Lu[1]

In this paper we mainly refine the recent entropy upper bound given by Ţăpuş and Popescu (2012). By using Jensen's inequality and some new inequalities with exponential functions and logarithmic functions, we obtain the stronger upper bound for entropy. At last we prove that the new upper bound is better than the previous one.

Keywords: Jensen's inequality, concave, entropy, new upper bound, refinement
MSC2010: 26B25; 26D15; 94A17.

1. Introduction

In information theory[2], if the probability distribution F is given by $P(X = i) = p_i,\ p_i > 0,\ i = 1, 2, \cdots, n,\ s.t.\ \sum_{i=1}^{n} p_i = 1$, then the (Shannon's) entropy is defined as $H(X) := \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$. The entropy reflects the expected value (or average value) of the information contained in each message. And it plays an important role in information science and applied mathematics. Some available bounds for the entropy can be seen in [1, 3, 4, 5, 6, 7, 8, 9, 10]. Recently, Simic improved the Jensen's inequality and obtained a new bound for the entropy with two given variables as follows[13]:

$$0 \le m(\mu, \nu) := \mu \log\left(\frac{2\mu}{\mu+\nu}\right) + \nu \log\left(\frac{2\nu}{\mu+\nu}\right) \le \log n - H(X),$$

where $\mu = \min_{1\le i\le n}\{p_i\}$ and $\nu = \max_{1\le i\le n}\{p_i\}$.

In 2012, Ţăpuş and Popescu [14] obtained a sharper entropy upper bound by using another refinement of Jensen's inequality based on Simic's work:

$$H(P) \le \log n - \max_{1\le \mu_1<\mu_2<\cdots<\mu_{n-1}\le n} \log\left[\left(\frac{n-1}{\sum_{i=1}^{n-1} p_{\mu_i}}\right)^{\sum_{i=1}^{n-1} p_{\mu_i}} \left(\prod_{i=1}^{n-1} p_{\mu_i}^{p_{\mu_i}}\right)\right]. \tag{1}$$

And soon Popescu et al. found a new upper bound for the entropy by a novel approach in modeling of big Data applications[11] and bounds for Kullback-Leibler divergence[12]. In this paper, we will obtain a more precise upper bound for entropy by using Jensen's inequality and some new inequalities with exponential functions and logarithmic functions.

2. Some preliminary results

In this paper, the term "log" refers to natural logarithm.

Theorem 2.1. *For $x > 0$,*

$$x - 1 - \log x \ge 1 + x(\mathrm{e}^{1-\frac{1}{x}} - 2). \tag{2}$$

The equality holds if and only if $x = 1$.

[1] Zhongnan University of Economics and Law, China, Email: lgxmath@zuel.edu.cn

Proof. Let $f(x)=x-1-\log x-\left[1+x(\mathrm{e}^{1-\frac{1}{x}}-2)\right]$. Direct computing yields

$$f'(x)=3-\frac{1}{x}-\left(1+\frac{1}{x}\right)\mathrm{e}^{1-\frac{1}{x}},$$
$$f''(x)=\frac{1}{x^3}\left(x-\mathrm{e}^{1-\frac{1}{x}}\right).$$

Using the standard inequality $\log x\ge 1-\frac{1}{x}$ with the necessary and sufficient condition $x=1$ for the equality, we can find that $x\ge \mathrm{e}^{1-\frac{1}{x}}$ and the equality holds if and only if $x=1$. Hence $f''(x)>0$ for $x\neq 1$. Since $f(1)=f'(1)=0$, we first have $f'(x)<0$ for $0<x<1$ and $f'(x)>0$ for $x>1$. Next we can obtain $f(x)>0$ for $0<x<1$ as well as $x>1$ and $f(x)=0$ for $x=1$. So the proof is complete. □

Theorem 2.2. *If* $\mathbf{a}=(a_1,a_2)$ *and* $\mathbf{q}=(q_1,q_2)$ *are two positive* 2*-tuples such that* $q_1+q_2=1$*, then*

$$\begin{aligned}\log(q_1a_1+q_2a_2)\ge&\frac{q_1a_1}{q_1a_1+q_2a_2}\left(\mathrm{e}^{1-\frac{q_1a_1+q_2a_2}{a_1}}-1\right)+\frac{q_2a_2}{q_1a_1+q_2a_2}\left(\mathrm{e}^{1-\frac{q_1a_1+q_2a_2}{a_2}}-1\right)\\&+(q_1\log a_1+q_2\log a_2).\end{aligned}\tag{3}$$

The equality holds if and only if $a_1=a_2$.

Proof. Substituting $x=\frac{a_i}{q_1a_1+q_2a_2}$ into (2) and multiplying by q_i for $i=1,2$, we have

$$\frac{q_ia_i}{q_1a_1+q_2a_2}-q_i-q_i\log\frac{a_i}{q_1a_1+q_2a_2}\ge q_i+\frac{q_ia_i}{q_1a_1+q_2a_2}\left(\mathrm{e}^{1-\frac{q_1a_1+q_2a_2}{a_i}}-2\right)$$

and the equality holds if and only if $a_1=a_2$ are equal. By using $q_1+q_2=1$ and $\frac{q_1a_1}{q_1a_1+q_2a_2}+\frac{q_2a_2}{q_1a_1+q_2a_2}=1$, after summing the two inequalities above we have

$$\begin{aligned}&-q_1\log\frac{a_1}{q_1a_1+q_2a_2}-q_2\log\frac{a_2}{q_1a_1+q_2a_2}\\\ge&\frac{q_1a_1}{q_1a_1+q_2a_2}\left(\mathrm{e}^{1-\frac{q_1a_1+q_2a_2}{a_1}}-1\right)+\frac{q_2a_2}{q_1a_1+q_2a_2}\left(\mathrm{e}^{1-\frac{q_1a_1+q_2a_2}{a_2}}-1\right).\end{aligned}$$

Therefore, the desired result follows and the equality holds if and only if $a_1=a_2$. □

Theorem 2.3. *Let* $f_\lambda(x):=\frac{x}{\lambda}\left(\mathrm{e}^{1-\frac{\lambda}{x}}-1\right)+\log x$, $\lambda>0$. *Then* f_λ *is a concave function on* $(0,+\infty)$.

Proof. Straightforward derivative shows

$$f''_\lambda(x)=\frac{1}{x^3}\left(\lambda\mathrm{e}^{1-\frac{\lambda}{x}}-x\right).$$

Observing the standard inequality $\log x\ge 1-\frac{1}{x}$, we have $\log\frac{x}{\lambda}\ge 1-\frac{\lambda}{x}$. Then the inequality $\frac{x}{\lambda}\ge \mathrm{e}^{1-\frac{\lambda}{x}}$ holds, or equivalently $\lambda\mathrm{e}^{1-\frac{\lambda}{x}}-x\le 0$. So the function $f_\lambda(x)$ is a concave function on $(0,+\infty)$. □

Next let $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ and $\mathbf{p}=(p_1,p_2,\cdots,p_n)$ are two positive n-tuple such that $\sum_{i=1}^n p_i=1$. Then we present two effective theorems.

Theorem 2.4. *If* f_λ *is defined as above,* P_n *is defined as* $P_n=\sum_{i=1}^n p_ix_i$*, and the notation* T_j *is defined as follows:*

$$T_j:=\max_{1\le\mu_1<\mu_2<\cdots<\mu_j\le n}\left[\left(\sum_{i=1}^j p_{\mu_i}\right)f_{P_n}\left(\frac{\sum_{i=1}^j p_{\mu_i}x_{\mu_i}}{\sum_{i=1}^j p_{\mu_i}}\right)-\sum_{i=1}^j p_{\mu_i}f_{P_n}(x_{\mu_i})\right],\tag{4}$$

where $j=2,\cdots,n-1$, then we have

$$0\le T_2\le T_3\le\cdots\le T_{n-1}.$$

Proof. Because $f_{P_n}(x)$ is concave on $(0,+\infty)$ by Theorem 2.3, using Jensen's inequality we can easily have $T_2 \geq 0$. Next we will show that for any $j \in \{2,\cdots,n-2\}$, $T_j \leq T_{j+1}$. Let us consider that the maximum of the expression

$$\left(\sum_{i=1}^{j} p_{\mu_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{j} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{j} p_{\mu_i}}\right) - \sum_{i=1}^{j} p_{\mu_i} f_{P_n}\left(x_{\mu_i}\right)$$

is obtained for $\mu_i = \eta_i$, $\eta_i \in \{1,2,\cdots,n\}$, $i = 1,2,\cdots,j$. Then it is enough to prove that

$$\begin{aligned}&\left(\sum_{i=1}^{j} p_{\eta_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{j} p_{\eta_i} x_{\eta_i}}{\sum_{i=1}^{j} p_{\eta_i}}\right) - \sum_{i=1}^{j} p_{\eta_i} f_{P_n}\left(x_{\eta_i}\right)\\ \leq &\left(\sum_{i=1}^{j+1} p_{\eta_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{j+1} p_{\eta_i} x_{\eta_i}}{\sum_{i=1}^{j+1} p_{\eta_i}}\right) - \sum_{i=1}^{j+1} p_{\eta_i} f_{P_n}\left(x_{\eta_i}\right)\end{aligned}$$

for any $\eta_{j+1} \in \{1,2,\cdots,n\} \setminus \{\eta_1,\cdots,\eta_j\}$. The above inequality is equivalent to

$$p_{\eta_{j+1}} f_{P_n}\left(x_{\eta_{j+1}}\right) + \left(\sum_{i=1}^{j} p_{\eta_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{j} p_{\eta_i} x_{\eta_i}}{\sum_{i=1}^{j} p_{\eta_i}}\right) \leq \left(\sum_{i=1}^{j+1} p_{\eta_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{j+1} p_{\eta_i} x_{\eta_i}}{\sum_{i=1}^{j+1} p_{\eta_i}}\right).$$

Multiplying by $\left(\sum_{i=1}^{j+1} p_{\eta_i}\right)^{-1}$, we have

$$\frac{p_{\eta_{j+1}}}{\sum_{i=1}^{j+1} p_{\eta_i}} f_{P_n}\left(x_{\eta_{j+1}}\right) + \frac{\sum_{i=1}^{j} p_{\eta_i}}{\sum_{i=1}^{j+1} p_{\eta_i}} f_{P_n}\left(\frac{\sum_{i=1}^{j} p_{\eta_i} x_{\eta_i}}{\sum_{i=1}^{j} p_{\eta_i}}\right) \leq f_{P_n}\left(\frac{\sum_{i=1}^{j+1} p_{\eta_i} x_{\eta_i}}{\sum_{i=1}^{j+1} p_{\eta_i}}\right).$$

This inequality follows from Jensen's inequality for the concave function $f_{P_n}(x)$. So we obtain the desired result. □

Theorem 2.5. *Let* $S = \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i \left(\mathrm{e}^{1-\frac{P_n}{x_i}} - 1\right)$, *then the following estimates hold*

$$S \leq S + T_2 \leq S + T_3 \leq \cdots \leq S + T_{n-1} \leq \log P_n - \sum_{i=1}^{n} p_i \log x_i. \tag{5}$$

Proof. By Theorem 2.4, we have

$$S \leq S + T_2 \leq S + T_3 \leq \cdots \leq S + T_{n-1}.$$

Next we prove the last inequality of (5). Choose arbitrary $x_{\mu_i} \in \{x_1, x_2, \cdots, x_n\}$ such that $1 \leq \mu_1 < \mu_2 < \cdots < \mu_{n-1} \leq n$ with corresponding multiplier $\{p_{\mu_1}, p_{\mu_2}, \cdots, p_{\mu_{n-1}}\}$, and let $x_{\mu_n} = \{x_1, x_2, \cdots, x_n\} \setminus \{x_{\mu_1}, x_{\mu_2}, \cdots, x_{\mu_{n-1}}\}$. Using the inequality (3) for $q_1 = p_{\mu_n}$,

$q_2 = \sum_{i=1}^{n-1} p_{\mu_i}$, $a_1 = x_{\mu_n}$, $a_2 = \frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}$, $q_1 a_1 + q_2 a_2 = P_n$, we have

$$\begin{aligned}
\log P_n &= \log\left(\sum_{i=1}^{n} p_i x_i\right) = \log\left(p_{\mu_n} x_{\mu_n} + \left(\sum_{i=1}^{n-1} p_{\mu_i}\right)\frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}\right)\\
&\geq \frac{1}{P_n}\left[p_{\mu_n} x_{\mu_n}\left(\mathrm{e}^{1-\frac{P_n}{x_{\mu_n}}} - 1\right) + \left(\sum_{i=1}^{n-1} p_{\mu_i}\right)\left(\frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}\right)\right.\\
&\quad \left.\cdot\left(\mathrm{e}^{1-\frac{P_n}{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i} / \sum_{i=1}^{n-1} p_{\mu_i}}} - 1\right)\right] + p_{\mu_n}\log x_{\mu_n} + \left(\sum_{i=1}^{n-1} p_{\mu_i}\right)\log\frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}\\
&= \frac{1}{P_n}\left(\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}\right)\left(\mathrm{e}^{1-\frac{P_n}{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i} / \sum_{i=1}^{n-1} p_{\mu_i}}} - 1\right) + \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\left(\mathrm{e}^{1-\frac{P_n}{x_i}} - 1\right)\\
&\quad - \frac{1}{P_n}\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}\left(\mathrm{e}^{1-\frac{P_n}{x_{\mu_i}}} - 1\right) + \sum_{i=1}^{n} p_i \log x_i - \sum_{i=1}^{n-1} p_{\mu_i}\log x_{\mu_i}\\
&\quad + \left(\sum_{i=1}^{n-1} p_{\mu_i}\right)\log\frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}\\
&= \sum_{i=1}^{n} p_i \log x_i + S + \left(\sum_{i=1}^{n-1} p_{\mu_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}\right) - \sum_{i=1}^{n-1} p_{\mu_i} f_{P_n}\left(x_{\mu_i}\right)
\end{aligned}$$

Because $\mu_i \in \{1, 2, \cdots, n\}$ are arbitrary, we have

$$\begin{aligned}
\log P_n &\geq \sum_{i=1}^{n} p_i \log x_i + S + \max_{1\leq\mu_1<\mu_2<\cdots<\mu_{n-1}\leq n}\left[\left(\sum_{i=1}^{n-1} p_{\mu_i}\right) f_{P_n}\left(\frac{\sum_{i=1}^{n-1} p_{\mu_i} x_{\mu_i}}{\sum_{i=1}^{n-1} p_{\mu_i}}\right)\right.\\
&\quad \left.-\sum_{i=1}^{n-1} p_{\mu_i} f_{P_n}\left(x_{\mu_i}\right)\right]\\
&= \sum_{i=1}^{n} p_i \log x_i + S + T_{n-1}.
\end{aligned}$$

Then the last inequality of (5) follows. □

3. The new upper bound for entropy

By using Theorem 2.5, we can improve the upper bound for entropy.

Theorem 3.1. *We have*

$$H(P) \leq \log n - \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{e}^{1-np_i} - 1\right) - \max_{1\leq\mu_1<\mu_2<\cdots<\mu_{n-1}\leq n}\{F(\mu) + G(\mu)\}, \tag{6}$$

where

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

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

and $\mu = (\mu_1, \mu_2, \cdots, \mu_n)$.

Proof. Applying the last inequality (5) with $x_i = 1/p_i$, after some calculations by using $A_n = n$ we can obtain the inequality (6). □

The following theorem can illustrate that our new upper bound (6) for entropy is better than previous bound (1) in [14].

Theorem 3.2.

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{e}^{1-np_i}-1\right)+\max_{1\le\mu_1<\mu_2<\cdots<\mu_{n-1}\le n}\{F(\mu)+G(\mu)\}\\ &\ge\max_{1\le\mu_1<\mu_2<\cdots<\mu_{n-1}\le n}\{F(\mu)\}.\end{aligned}\tag{7}$$

Proof. Let us consider that the maximum of the right-hand side of the inequality (7) is obtained for $\mu_i = \eta_i$, $\eta_i \in \{1, 2, \cdots, n\}$, $i = 1, 2, \cdots, n-1$, and let $\eta_n = \{1, 2, \cdots, n\} \setminus \{\eta_1, \cdots, \eta_{n-1}\}$. For $\eta = (\eta_1, \eta_2, \cdots, \eta_n)$ we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{e}^{1-np_i}-1\right)+\max_{1\le\mu_1<\mu_2<\cdots<\mu_{n-1}\le n}\{F(\mu)+G(\mu)\}\\ &-\max_{1\le\mu_1<\mu_2<\cdots<\mu_{n-1}\le n}\{F(\mu)\}\\ \ge&\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{e}^{1-np_i}-1\right)+G(\eta)\\ =&\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{e}^{1-np_i}-1\right)+\frac{n-1}{n}\left(\mathrm{e}^{1-\frac{n}{n-1}\sum_{i=1}^{n-1}p_{\eta_i}}-1\right)-\frac{1}{n}\sum_{i=1}^{n-1}\left(\mathrm{e}^{1-np_{\eta_i}}-1\right)\\ =&\frac{1}{n}\left(\mathrm{e}^{1-np_{\eta_n}}-1\right)+\frac{n-1}{n}\left(\mathrm{e}^{1-\frac{n}{n-1}\sum_{i=1}^{n-1}p_{\eta_i}}-1\right)\end{aligned}$$

Let $\psi(x) = \mathrm{e}^{1-x} - 1$. We can easily obtain $\psi(x)$ is convex for $x > 0$ by the second derivative $\psi''(x) = \mathrm{e}^{1-x} > 0$. Using Jensen's inequality we have

$$\begin{aligned}&\frac{1}{n}\left(\mathrm{e}^{1-np_{\eta_n}}-1\right)+\frac{n-1}{n}\left(\mathrm{e}^{1-\frac{n}{n-1}\sum_{i=1}^{n-1}p_{\eta_i}}-1\right)\\ &\ge\mathrm{e}^{1-\left(\frac{1}{n}np_{\eta_n}+\frac{n-1}{n}\frac{n}{n-1}\sum_{i=1}^{n-1}p_{\eta_i}\right)}-1=\mathrm{e}^{1-\sum_{i=1}^{n}p_i}-1=0\end{aligned}$$

So we obtain the desired result. □

4. Acknowledgement

The author would like to thank the editor and referees for their helpful suggestions and comments on the manuscript. This manuscript is supported by Hubei Provincial Natural Science Foundation of China (2017CFB030).

REFERENCES

[1] *I. Budimir I, S. S. Dragomir and J. Pečarić*, Further reverse results for Jensen's discrete inequality and applications in information theory, J. Inequal. Pure Appl. Math. **2**(2001), No. 1, Art. 5.

[2] *T. M. Cover and J. A. Thomas*, Elements of Information Theory, John Wiley and Sons, 2006.

[3] *S. S. Dragomir and C. J. Goh*, A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory, Math. Comput. Model., **24**(1996), No. 2, 1-11.

[4] *S.S. Dragomir and C. J. Goh*, Some bounds on entropy measures in information theory, Appl. Math. Lett., **10**(1997), No. 3, 23-28.

[5] *S.S. Dragomir, N. M. Dragomir and K. Pranesh*, An inequality for logarithms and applications in information theory, Comput. Math. Appl., **38**(1999), No. 2, 11-17.

[6] *S. S. Dragomir, M. L. Scholz and J. Sunde*, Some upper bounds for relative entropy and applications, Comput. Math. Appl., **39**(2000), No. 9, 91-100.

[7] *S.S. Dragomir*, An inequality for logarithmic mapping and applications for the shannon entropy, Comput. Math. Appl., **46**(2003), No. 8-9, 1273-1279.

[8] *M. Matić, C. E. M. Pearce and J. Pečarić*, Improvements of some bounds on entropy measures in information theory, Math. Inequal. Appl. **1**(1998), No. 2, 295C304.

[9] *M. Matić, C. E. M. Pearce and J. Pečarić*, Further improvements of some bounds on entropy measure in information theory, Math. Inequal. Appl, **2**(1999), No. 4, 599-611.

[10] *M. Matić, C. E. M. Pearce and J. Pečarić*, Refinements of some bounds in information theory, ANZIAM J., **42**(2001), No. 3, 387-398.

[11] *P. G. Popescu, E. I. Sluşanschi, V. Iancu and F. Pop*, A new upper bound for Shannon entropy. A novel approach in modeling of Big Data applications, Concurr. Computat.: Pract. Exper. **28**(2016), No. 2, 351-359.

[12] *P. G. Popescu, S. S. Dragomir, E. I. Sluşanschi and O. N. Stănăşilă*, Bounds for Kullback-Leibler divergence, Electron. J. Differ. Eq. **237**(2016), 1-6.

[13] *S. Simic*, Jensen's inequality and new entropy bounds, Appl. Math. Lett. **22**(2009), No. 8, 1262-1265.

[14] *N. Ţăpuş and P. G. Popescu*, A new entropy upper bound, Appl. Math. Lett. **25**(2012), No. 11, 1887-1890.