Electronic Journal of Differential Equations, Vol. 2016 (2016), No. 237, pp. 1–6.
ISSN: 1072-6691. URL: http://ejde.math.txstate.edu or http://ejde.math.unt.edu

BOUNDS FOR KULLBACK-LEIBLER DIVERGENCE

PANTELIMON G. POPESCU, SEVER S. DRAGOMIR,
EMIL I. SLUŞANSCHI, OCTAVIAN N. STĂNĂŞILĂ

Abstract. Entropy, conditional entropy and mutual information for discrete-valued random variables play important roles in the information theory. The purpose of this paper is to present new bounds for relative entropy $D(p||q)$ of two probability distributions and then to apply them to simple entropy and mutual information. The relative entropy upper bound obtained is a refinement of a bound previously presented into literature.

1. Introduction

The relative entropy $D(p||q)$ (see [1],[2]) is the measure of distance between two distributions. It can also be expressed like a measure of the inefficiency of assuming that the distribution is q when the true distribution is p.

Definition 1.1. The relative entropy, of the Kullback-Leibler distance, between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$D(p||q) := \sum_{x\in X} p(x)\ln\big(\frac{p(x)}{q(x)}\big) = E_p \ln\big(\frac{p}{q}\big),$$

where $\ln(\cdot)$ is the natural logarithm.

A fundamental property of the relative entropy is the following.

Theorem 1.2. *Let $p(x), q(x), x \in X$ be two probability mass functions. Then*

$$D(p||q) \geq 0,$$

with equality if and only if $p(x) = q(x)$ for all $x \in X$.

Obtaining a lower bound, the above fundamental inequality can be improved as follows (see [2]).

Theorem 1.3. *Let $p(x), q(x)$ be two probability mass functions for $x \in X$. Then*

$$D(p||q) \geq \frac{1}{2}\Big(\sum_{x\in X} |p(x) - q(x)|\Big)^2.$$

As an upper bound for relative entropy, we have taken into consideration the result presented by Dragomir et al. [3, Theorem 1]. Other bounds can also be found in [4, 5, 6, 7].

2010 *Mathematics Subject Classification.* 26B25, 94A17.
Key words and phrases. Entropy; bounds; refinements; generalization.

Submitted June 19, 2016. Published August 30, 2016.

Theorem 1.4 ([3]). *Let $p(x), q(x) > 0, x \in X$ be two probability mass functions. Then*

$$D(p||q) \leq \sum_{x \in X} \frac{p^2(x)}{q(x)} - 1 = \frac{1}{2} \sum_{x,y \in X} p(x)q(x)\Big(\frac{p(x)}{q(x)} - \frac{p(y)}{q(y)}\Big)\Big(\frac{q(y)}{p(y)} - \frac{q(x)}{p(x)}\Big),$$

with equality if and only if $p(x) = q(x)$ for all $x \in X$.

2. A new inequality for strictly convex functions

First we need a generalization of the classical Lagrange's theorem.

Theorem 2.1. *Let n be a natural number, and $f : [a, b] \to \mathbb{R}$ a continuous function on $[a, b]$ and derivable on (a, b), with $b > a > 0$. Then exists distinct $c_1, c_2, \dots, c_n \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} = \frac{f'(c_1) + f'(c_2) + \dots + f'(c_n)}{n}.$$

Proof. First we split the interval $[a, b]$ into n equal subintervals, as $[a, x_1]$, $[x_1, x_2]$, $\dots$, $[x_{n-1}, b]$, with

$$x_1 - a = x_2 - x_1 = \dots = b - x_{n-1} = \frac{b-a}{n}.$$

Now applying the Lagrange Theorem for each interval, we obtain that exists $c_1 \in [a, x_1]$, $c_2 \in [x_1, x_2], \dots, c_n \in [x_{n-1}, b]$ such that

$$\frac{f(x_1) - f(a)}{\frac{b-a}{n}} = f'(c_1), \quad \frac{f(x_2) - f(x_1)}{\frac{b-a}{n}} = f'(c_2), \dots, \frac{f(b) - f(x_{n-1})}{\frac{b-a}{n}} = f'(c_n).$$

Summing the above equalities, yields that

$$\frac{f(b) - f(a)}{b - a} = \frac{f'(c_1) + f'(c_2) + \dots + f'(c_n)}{n}.$$

the proof is complete. □

Applying the condition of strictly convex functions to the function f, we obtain the following result.

Theorem 2.2. *Let n be a natural number, and $f : [a, b] \to \mathbb{R}$ a continuous function on $[a, b]$, diffferentiable on (a, b) and strictly convex, with $b > a > 0$. Then*

$$f'(a) + \sum_{i=1}^{n-1} f'\Big(a + i\frac{b-a}{n}\Big) < n\frac{f(b) - f(a)}{b - a} < \sum_{i=1}^{n-1} f'\Big(a + i\frac{b-a}{n}\Big) + f'(b).$$

Proof. As f is a strictly convex function implies that f' is an increasing function, so because from the above theorem exist $c_1 \in [a, x_1]$, $c_2 \in [x_1, x_2]$, $\dots$, $c_n \in [x_{n-1}, b]$, with $x_1 - a = x_2 - x_1 = \dots = b - x_{n-1} = \frac{b-a}{n}$. This implies

$$f'(a) < f'(c_1) < f'(x_1), f'(x_1) < f'(c_2) < f'(x_2), \dots, f'(x_{n-1}) < f'(c_n) < f'(b)$$

and considering the result of the previous theorem and summating, we get the wanted result. □

Remark 2.3. It is easy to see that for any positive natural number n

$$nf'(a) \le f'(a) + \sum_{i=1}^{n-1} f'\Big(a + i\frac{b-a}{n}\Big)$$

and

$$\sum_{i=1}^{n-1} f'\Big(a + i\frac{b-a}{n}\Big) + f'(b) \le nf'(b),$$

because $a < a + i\frac{b-a}{n}$ and $a + i\frac{b-a}{n} < b$, for $i = 1, 2, \dots, n-1$.

3. New bounds for relative entropy

To present a general inequality for $-\ln x$ we start with a helpful result, which can be deducted by simple calculus.

Lemma 3.1. *Let a, b, t, T be real numbers with $b \neq 0$ and $T > t > 0$. Then the following two inequalities are equivalent*

$$t < \frac{a}{b} < T$$

and

$$b\frac{T + t - \frac{b}{|b|}(T-t)}{2} < a < b\frac{T + t + \frac{b}{|b|}(T-t)}{2}.$$

Now, applying Theorem 2.2 to the function $-\ln x$ and taking into consideration the previous Lemma, yields the following result.

Corollary 3.2. *Let $a, b > 0$, with $m = \min\{a, b\}$, $M = \max\{a, b\}$ and $n \ge 1$ a natural number, then*

$$\begin{aligned}
(a-b)\frac{1}{a} &\le (a-b)\Big(\sum_{i=1}^{n-1}\frac{1}{nm + i(M-m)} + \frac{m+M+b-a}{2nmM}\Big) \quad \le \ln a - \ln b \\
&\le (a-b)\Big(\sum_{i=1}^{n-1}\frac{1}{nm + i(M-m)} + \frac{m+M+a-b}{2nmM}\Big) \\
&\le (a-b)\frac{1}{b},
\end{aligned}$$

with equality holding for $a = b$.

Proof. If $a = b$ then the inequality is obvious, so if $a \neq b$, by applying Theorem 2.2 to the function $-\ln x$ defined on the interval $I = [m, M]$ and taking cont that $\frac{f(M)-f(m)}{M-m} = \frac{f(b)-f(a)}{b-a}$ we obtain

$$-\frac{1}{nm} - \sum_{i=1}^{n-1}\frac{1}{nm + i(M-m)} < \frac{\ln a - \ln b}{b-a} < -\sum_{i=1}^{n-1}\frac{1}{nm + i(M-m)} - \frac{1}{nM}\,.$$

The equivalence from Lemma 3.1, yields

$$\begin{aligned}
&(a-b)\frac{2\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M}{nmM} - \frac{b-a}{M-m}\frac{m-M}{nmM}}{2} \\
&< \ln a - \ln b \\
&< (a-b)\frac{2\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M}{nmM} + \frac{b-a}{M-m}\frac{m-M}{nmM}}{2},
\end{aligned}$$

which is equivalent to

$$\begin{aligned}&(a-b)\Big(\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)}+\frac{m+M+b-a}{2nmM}\Big)\\&<\ln a-\ln b\\&<(a-b)\Big(\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)}+\frac{m+M+a-b}{2nmM}\Big)\end{aligned}$$

and by comparing with the previous remark we obtain the wanted result. □

We present the main result of this article, namely, new bounds for the relative entropy, where we have considered two probability mass function $p(x)$ and $q(x), x \in X$.

Theorem 3.3. *Let $p(x), q(x) > 0, x \in X$ be two probability mass functions, with $m(x) = \min\{p(x), q(x)\}$ and $M(x) = \max\{p(x), q(x)\}, x \in X$. If $r(x) = p(x)-q(x)$, then*

$$\begin{aligned}&\sum_{x\in X}p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\Big)\\&\geq D(p||q)\\&\geq\sum_{x\in X}p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\Big),\end{aligned}$$

with equality if and only if $p(x) = q(x)$ for all $x \in X$.

Proof. Setting $a = q(x)$ and $b = p(x)$ in Corollary 3.2 we obtain

$$\begin{aligned}&-r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\Big)\\&\leq\ln q(x)-\ln p(x)\\&\leq -r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\Big),\end{aligned}$$

and multiplying by $-p(x)$ yields

$$\begin{aligned}&p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\Big)\\&\geq p(x)\ln\frac{p(x)}{q(x)}\\&\geq p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\Big),\end{aligned}$$

from which summing over $x \in X$ we get the wanted result. □

Remark 3.4. From Corollary 3.2 and Theorem 3.3 we deduce that

$$\sum_{x\in X}\frac{p^2(x)}{q(x)}-1$$

$$
\begin{aligned}
&\geq \sum_{x\in X} p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\Big)\\
&\geq D(p||q),
\end{aligned}
$$

which leads us to the conclusion that the upper bound of relative entropy provided by Theorem 3.3 is stronger that the one from [3].

Furthermore we present new bounds for entropy and mutual information.

Corollary 3.5. *Let X be a random variable whose range has $|X|$ elements and has the probability mass function $p(x) > 0$, with $m(x) = \min\{p(x), 1/|X|\}$ and $M(x) = \max\{p(x), 1/|X|\}, x \in X$. If $r(x) = p(x) - 1/|X|$, then*

$$
\begin{aligned}
&\sum_{x\in X} p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\Big)\\
&\geq \ln|X| - H(X)\\
&\geq \sum_{x\in X} p(x)r(x)\Big(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\Big).
\end{aligned}
$$

The equality holds if and only if $p(x) = 1/|X|$.

Proof. It follows from Theorem 3.3 applied for $D(p||q)$, where $p(x) = p(x)$ and $q(x) = 1/|X|)$, i.e. $D(p(x)||1/|X|)$. □

Corollary 3.6. *Let X, Y be two random variables with a joint probability mass function $p(x, y)$ and marginal probability mass function $p(x)$ and $p(y)$, with $p(x, y)$, $p(x)$, $p(y) > 0$, $x \in X, y \in Y$ and $m_{x,y} = \min\{p(x, y), p(x)p(y)\}$ and $M_{x,y} = \max\{p(x, y), p(x)p(y)\}, x \in X, y \in Y$. If $r_{x,y} = p(x, y) - p(x)p(y)$, then*

$$
\begin{aligned}
&\sum_{(x,y)\in X\times Y} p(x,y)r_{x,y}\Big(\sum_{i=1}^{n-1}\frac{1}{nm_{x,y}+i(M_{x,y}-m_{x,y})}+\frac{m_{x,y}+M_{x,y}+r_{x,y}}{2nm_{x,y}M_{x,y}}\Big)\\
&\geq I(X;Y)\\
&\geq \sum_{(x,y)\in X\times Y} p(x,y)r_{x,y}\Big(\sum_{i=1}^{n-1}\frac{1}{nm_{x,y}+i(M_{x,y}-m_{x,y})}+\frac{m_{x,y}+M_{x,y}-r_{x,y}}{2nm_{x,y}M_{x,y}}\Big),
\end{aligned}
$$

The equality holds if and only if X and Y are independent.

Proof. It follows from Theorem 3.3 applied for $D(p||q)$, where $p(x) = p(x, y)$ and $q(x) = p(x)p(y)$, i.e. $D(p(x, y)||p(x)p(y))$ and where $m(x) = m_{x,y}, M(x) = M_{x,y}, r(x) = r_{x,y}$. □

References

[1] R. Ash; *Information Theory*, Interscience, New York, 1965.

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

[3] S. S. Dragomir, M. L. Scholz, J. Sunde; *Some upper bounds for relative entropy and applications*, Computers and Mathematics with Applications, 39 (2000), 91-100.

[4] S. S. Dragomir, C. J. Goh; *A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory*, Math. Comput. Modelling, 24 (1996), 1-11.

[5] M. Matic, C. E. M. Pearce, J. Pecaric; *Refinements of some bounds in information theory*, ANZIAM J., 42 (2001), 387-398.
[6] S. Simic; *Jensen's inequality and new entropy bounds*, Applied Mathematics Letters, 22(2009), 1262-1265.
[7] N. Tapus, P. G. Popescu; *New entropy upper bound*, Applied Mathematics Letters, 25 no. 11 (2014), 1887–1890.

Pantelimon George Popescu
Computer Science and Engineering Departament, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independenţei 313, 060042, Bucharest (6), Romania
E-mail address: pgpopescu@yahoo.com, Phone +40741533097, Fax +40214029333

Sever Silvestru Dragomir
College of Engineering and Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia
E-mail address: sever.dragomir@vu.edu.au, Phone +61 3 9919 4437, Fax +61 3 9919 4050

Emil Ioan Sluşanschi
Computer Science and Engineering Departament, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independenţei 313, 060042, Bucharest (6), Romania
E-mail address: emil.slusanschi@cs.pub.ro, Phone +40741533097, Fax +40214029333

Octavian Nicolae Stănăşilă
Computer Science and Engineering Departament, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independenţei 313, 060042, Bucharest (6), Romania
E-mail address: ostanasila@hotmail.com, Phone +40741533097, Fax +40214029333