Received 12/02/15

New Bounds for Kullback-Leibler Divergence based on a Lagrange Theorem

S.S. Dragomir[a,*], P.G.Popescu[b], N.Ţăpuş[b], E.I.Sluşanschi[b]

[a] *College of Engineering and Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia, Phone: +61 3 9919 4437, Fax: +61 3 9919 4050*
[b] *Computer Science and Engineering Departament, Faculty of Automatic Control and Computers, University "Politehnica" of Bucharest, Splaiul Independenţei 313, 060042, Bucharest (6), Romania, Phone: +40741533097, Fax: +40214029333*

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. For this purpose we develop a generalization for the classical Lagrange's Theorem and we apply it to refine the wanted result.

Keywords: Lagrange's theorem, entropy, bounds, refinements, generalizations

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

$$D(p||q) := \sum_{x\in X} p(x)\log\left(\frac{p(x)}{q(x)}\right) = E_p \log\left(\frac{p}{q}\right),$$

where log *is the natural logarithm.*

*Corresponding author
Email addresses: sever.dragomir@vu.edu.au (S.S. Dragomir), pgpopescu@yahoo.com (P.G.Popescu), ntapus@cs.pub.ro (N.Ţăpuş), emil.slusanschi@cs.pub.ro (E.I.Sluşanschi)

A fundamental property of the relative entropy is the following

Theorem 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 $p(x) = q(x), \forall x \in X$.

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

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

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

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

Theorem 4. **(S.S. Dragomir et al.)** *Let $p(x), q(x) > 0, x \in X$ be two probability mass functions. Then*

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

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

Furthermore, in order to refine the previous inequality, we continue presenting Lagrange's Theorem, for which we will express a generalization that will help us tighten the result.

Theorem 5. **(Lagrange)** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$ and derivable on (a, b), with $b > a > 0$. Then exists $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

2. Generalization of Lagrange's Theorem and a New Inequality for Strictly Convex Functions

We continue with a generalization of Lagrange's theorem. A similar previous result can be found in [8], as follows

Theorem 6. *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 $c_1, c_2, ..., c_n \in (a, b)$ such that*

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

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

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

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

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

Continue with summating the above equalities, yields that

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

and we are done. □

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

Theorem 7. *Let n be a natural number, and $f : [a, b] \to \mathbb{R}$ a continuous function on $[a, b]$, derivable on (a, b) and strictly convex, with $b > a > 0$. Then*

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

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

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

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

We continue with the following remark

Remark 8. *It is easy to see that for any natural positive number n*

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

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

3. New Bounds for Relative Entropy

In order to present a general inequality for $-\log x$ we start with a helpful result, which can be deducted by simple calculus, as follows

Lemma 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 7 to the function $-\log x$ and taking into consideration the previous Lemma, yields

Corollary 9. *Let $a, b > 0$, with $m = \min\{a, b\}, M = \max\{a, b\}$ and $n \geq 1$ a natural number, then*

$$(a-b)\frac{1}{a} \leq (a-b)\left(\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M+b-a}{2nmM}\right) \leq \log a - \log b$$

$$\leq (a-b)\left(\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M+a-b}{2nmM}\right) \leq (a-b)\frac{1}{b},$$

with equality holding for $a = b$.

Proof. If $a = b$ then the inequality is obvious, so if $a \neq b$, by applying Theorem 7 to the function $-\log x$ 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 get

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

and now taking into consideration the equivalence from Lemma 1, yields

$$(a-b)\frac{2\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M}{nmM} - \frac{b-a}{M-m}\frac{m-M}{nmM}}{2} < \log a - \log b$$

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

which is equivalent with

$$(a-b)\left(\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M+b-a}{2nmM}\right) < \log a - \log b$$

$$< (a-b)\left(\sum_{i=1}^{n-1}\frac{1}{nm+i(M-m)} + \frac{m+M+a-b}{2nmM}\right)$$

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

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

Theorem 10. *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*

$$\sum_{x\in X} p(x)r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))} + \frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\right)$$

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

$$\geq \sum_{x\in X} p(x)r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))} + \frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\right),$$

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

Proof. Setting $a = q(x)$ and $b = p(x)$ in Corollary 9 we obtain

$$-r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\right)$$
$$\leq \log q(x) - \log p(x)$$
$$\leq -r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\right),$$

and multiplying by $-p(x)$ yields

$$p(x)r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\right)$$
$$\geq p(x)\log\frac{p(x)}{q(x)}$$
$$\geq p(x)r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\right),$$

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

We continue with the following remark

Remark 11. *From Corollary 9 and Theorem 10 we can deduce that*

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

which leads us to the conclusion that the upper bound of relative entropy provided by Theorem 10 is stronger that the one from [3].

Furthermore we present new bounds for entropy and mutual information.

Corollary 12. *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*

$$\sum_{x\in X}p(x)r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)+r(x)}{2nm(x)M(x)}\right)$$
$$\geq \log|X| - H(X)$$
$$\geq \sum_{x\in X}p(x)r(x)\left(\sum_{i=1}^{n-1}\frac{1}{nm(x)+i(M(x)-m(x))}+\frac{m(x)+M(x)-r(x)}{2nm(x)M(x)}\right).$$

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

Proof. It follows from Theorem 10 applied for $D(p||q)$, where now $p(x) = p(x)$ and $q(x) = 1/|X|)$, i.e. $D(p(x)||1/|X|)$. □

Corollary 13. *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*

$$\sum_{(x,y)\in X\times Y} p(x,y) r_{x,y} \left(\sum_{i=1}^{n-1} \frac{1}{nm_{x,y} + i(M_{x,y} - m_{x,y})} + \frac{m_{x,y} + M_{x,y} + r_{x,y}}{2nm_{x,y}M_{x,y}} \right)$$

$$\geq I(X;Y)$$

$$\geq \sum_{(x,y)\in X\times Y} p(x,y) r_{x,y} \left(\sum_{i=1}^{n-1} \frac{1}{nm_{x,y} + i(M_{x,y} - m_{x,y})} + \frac{m_{x,y} + M_{x,y} - r_{x,y}}{2nm_{x,y}M_{x,y}} \right),$$

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

Proof. It follows from Theorem 10 applied for $D(p||q)$, where now $p(x) = p(x,y)$ and $q(x) = p(x)p(y)$, i.e. $D(p(x,y)||p(x)p(y))$ and where, this time $m(x) = m_{x,y}, M(x) = M_{x,y}, r(x) = r_{x,y}$. □

Acknowledgements

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398.

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

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

[3] S. S. Dragomir, M. L. Scholz and J. Sunde, Some Upper Bounds for Relative Entropy and Applications, Computers and Mathematics with Applications, 39(2000), 91-100.

[4] 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. Modelling, 24(1996), 1-11.

[5] M. Matic, C. E. M. Pearce and 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 and P.G. Popescu, New Entropy Upper Bound, Applied Mathematics Letters, 25 no. 11(2014), 1887–1890.

[8] P. G. Popescu, O extindere a teoremei lui Lagrange, Revista Arhimede, 1-2(2004), 5-8.