# *Some Inequalities for the Relative Entropy and Applications*

# SOME INEQUALITIES FOR THE RELATIVE ENTROPY AND APPLICATIONS

S.S. DRAGOMIR

Abstract. Some new inequalities for the relative entropy and applications are given.

## 1. Introduction

The *relative entropy* is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(p\|q)$ is a measure of the inefficiency of assuming that the distribution is $q$ when the true distribution is $p$. For example, if we knew the true distribution of the random variable, then we could construct a code with average description length $H(p)$. If, instead, we used the code for a distribution $q$, we would need $H(p)+D(p\|q)$ bits on the average to describe the random variable [1, p. 18].

In what follows, unless we will specify, log will denote $\log_b$ where $b$ is a given base greater than 1.

**Definition 1.** *The relative entropy or Kullback-Leibler distance between two probability mass functions $p(x)$ and $q(x)$ is defined by*

$$
\begin{aligned}
D(p\|q) \quad : &= \sum_{x\in\mathcal{X}} p(x)\log\left(\frac{p(x)}{q(x)}\right) \\
&= E_p \log\left(\frac{p(X)}{q(X)}\right).
\end{aligned} \tag{1.1}
$$

In the above definition, we use the convention (based on continuity arguments) that $0\log\left(\frac{0}{q}\right)=0$ and $p\log\left(\frac{p}{0}\right)=\infty$.

It is well-known that relative entropy is always non-negative and is zero if and only if $p=q$. However, it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality.

The following theorem is of fundamental importance [1, p. 26].

**Theorem 1.** *(Information Inequality) Let $p(x)$, $q(x)\in\mathcal{X}$, be two probability mass functions. Then*

$$D(p\|q)\geq 0 \tag{1.2}$$

*with equality if and only if*

$$p(x)=q(x) \quad \textit{for all } x\in\mathcal{X}. \tag{1.3}$$

---

*Date*: January , 1998.

1991 *Mathematics Subject Classification.* Primary 26D15; Secondary 94Xxx.

*Key words and phrases.* Relative Entropy, Shannon Entropy, Mutual Information.

Actually, the inequality (1.2) can be improved as follows (see , [1, p. 300]):

**Theorem 2.** *Let $p, q$ be as above. Then*

$$D(p\|q) \geq \frac{1}{2\ln b}\|p-q\|_1^2 \tag{1.4}$$

*where $\|p-q\|_1 = \sum_{x\in\mathcal{X}} |p(x)-q(x)|$ is the usual 1-norm of $p-q$. The equality holds iff $p=q$.*

We remark that the argument of (1.4) is not based on the convexity of the map $-\log(.)$.

To evaluate the relative entropy $D(p\|q)$ it would be interesting to establish some upper bounds.

Before we do this, let us recall some other important concepts in Information Theory.

We introduce *mutual information*, which is a measure of the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one random variable due to the knowledge of the other [1, p. 18].

**Definition 2.** *Consider two random variables $X$ and $Y$ with a joint probability mass function $p(x,y)$ and marginal probability mass function $p(x)$ and $q(y)$. The mutual information is the relative entropy between the joint distribution and the product distribution, i.e.*

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left(\frac{p(x,y)}{p(x)q(y)}\right) \\ &= D\left(p(x,y)\ \|\ p(x)q(y)\right) \\ &= E_{p(x,y)}\log\left(\frac{p(X,Y)}{p(X)q(Y)}\right). \end{aligned} \tag{1.5}$$

The following corollary of Theorem 1 holds [1, p. 27].

**Corollary 1.** *(Non-negativity of mutual information): For any two random variables, $X$, $Y$ we have*

$$I(X;Y) \geq 0 \tag{1.6}$$

*with equality if and only if $X$ and $Y$ are independent.*

An improvement of this result via Theorem 2 is as follows

**Corollary 2.** *For any two random variables, $X$, $Y$ we have*

$$I(X;Y) \geq \frac{1}{2\ln b}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} |p(x,y)-p(x)q(y)|\right]^2 \geq 0 \tag{1.7}$$

*with equality if and only if $X$ and $Y$ are independent.*

Now, let $u(x) = \frac{1}{|\mathcal{X}|}$ be the uniform probability mass function on $\mathcal{X}$ and let $p(x)$ be the probability mass function for $\mathbf{X}$.

It is well-known that [1, p. 27]

$$D(p\|u) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{u(x)} = \log|\mathcal{X}| - H(X). \tag{1.8}$$

The following corollary of Theorem 1 is important [1, p. 27].

**Corollary 3.** *Let $X$ be a random variable and $|\mathcal{X}|$ denotes the number of elements in the range of $X$. Then*

$$H(X) \le \log|\mathcal{X}| \tag{1.9}$$

*with equality if and only if $X$ has a uniform distribution over $\mathcal{X}$.*

Using Theorem 2 we also can state

**Corollary 4.** *Let $X$ be as above. Then*

$$\log|\mathcal{X}| - H(X) \ge \frac{1}{2\ln b}\left[\sum_{x\in\mathcal{X}}\left|p(x) - \frac{1}{|\mathcal{X}|}\right|\right]^2 \ge 0. \tag{1.10}$$

*The equality holds iff $p$ is uniformly distributed on $\mathcal{X}$ .*

In the recent paper [2], the authors proved between other the following upper bound for the relative entropy and employed it in Coding Theory in connection to Noiseless Coding Theorem:

**Theorem 3.** *Under the above assumptions for $p(x)$ and $q(x)$ we have the inequality*

$$\frac{1}{\ln b}\left[\sum_{x\in\mathcal{X}}\frac{p^2(x)}{q(x)} - 1\right] \ge D(p\|q) \tag{1.11}$$

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

The following upper bound for the mutual information holds

**Corollary 5.** *For any two random variables, $X$, $Y$ we have*

$$\frac{1}{\ln b}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\frac{p^2(x,y)}{p(x)q(y)} - 1\right] \ge I(X;Y)$$

*with equality iff $X$ and $Y$ are independent.*

Finally, we note that the following upper bound for the difference $\log|\mathcal{X}| - H(X)$ is valid

**Corollary 6.** *We have*

$$\frac{1}{\ln b}\left[|\mathcal{X}|\sum_{x\in\mathcal{X}} p^2(x) - 1\right] \ge \log|\mathcal{X}| - H(X)$$

*with equality if and only if $p$ is uniformly distributed on $\mathcal{X}$ .*

The main aim of this paper is to point out some refinements of (1.2) and (1.11) and to apply them for the entropy mapping and for mutual information .

## 2. An Inequality for Relative Entropy

The following result holds

**Theorem 4.** *Let $p(x), q(x), x \in \mathcal{X}$ be two probability mass functions. Then*

$$\frac{1}{\ln b}\left[\sum_{x\in\mathcal{X}}\frac{p^2(x)}{q(x)}-1\right] \geq \frac{1}{\ln b}\left\{\sum_{x\in\mathcal{X}}\left[\left(\frac{p(x)}{q(x)}\right)^{p(x)}-1\right]\right\} \tag{2.1}$$

$$\geq D(p\|q) \geq \frac{1}{\ln b}\left\{\sum_{x\in\mathcal{X}}\left[1-\left(\frac{q(x)}{p(x)}\right)^{p(x)}\right]\right\} \geq 0.$$

*The equality holds in all inequalities if and only if $p(x) = q(x)$ for all $x \in \mathcal{X}$ .*

**Proof**. We know that for every differentiable strictly convex mapping $f : I \to R$, we have the double inequality

$$f'(x)(x-y) \geq f(x) - f(y) \geq f'(y)(x-y)$$

for all $x, y \in I$. The equality holds iff $x = y$.

Now, if we apply this inequality to the strictly convex mapping $-\ln(\cdot)$ on the interval $(0, \infty)$ we deduce

$$\frac{1}{y}(x-y) \geq \ln x - \ln y \geq \frac{1}{x}(x-y)$$

for all $x, y > 0$, with equality iff $x = y$.

Now, choose $x = a^a$, $y = b^a$ in the above inequality to deduce

$$\left(\frac{a}{b}\right)^a - 1 \geq a\ln a - a\ln b \geq 1 - \left(\frac{b}{a}\right)^a \tag{2.2}$$

with equality iff $a = b$.

If in (2.2) we put $a = p(x)$, $b = q(x)$ , then we obtain

$$\left(\frac{p(x)}{q(x)}\right)^{p(x)} - 1 \geq p(x)\ln p(x) - p(x)\ln q(x) \geq 1 - \left(\frac{q(x)}{q(x)}\right)^{p(x)},$$

with equality iff $p(x) = q(x)$, $x \in \mathcal{X}$ .

Summing over $x \in \mathcal{X}$ we get

$$\sum_{x\in\mathcal{X}}\left[\left(\frac{p(x)}{q(x)}\right)^{p(x)}-1\right] \geq D_e(p\|q) \geq \sum_{x\in\mathcal{X}}\left[1-\left(\frac{q(x)}{p(x)}\right)^{p(x)}\right]$$

with equality iff $p(x) = q(x)$, $x \in \mathcal{X}$ .

Now, let observe that

$$\left(\frac{q(x)}{p(x)}\right)^{p(x)} = q(x)\left(\frac{1}{p(x)}\right)^{p(x)}\left(\frac{1}{q(x)}\right)^{1-p(x)}, x \in \mathcal{X}.$$

Using the *weighted arithmetic mean-geometric mean inequality, i.e.,* we recall it

$$a^t b^{1-t} \leq ta + (1-t)b, \quad a, b > 0, t \in (0,1)$$

with equality iff $a = b$, for the choices

$$a = \frac{1}{p(x)}, b = \frac{1}{q(x)}, t = p(x), x \in \mathcal{X}$$

we get

$$\left(\frac{1}{p(x)}\right)^{p(x)}\left(\frac{1}{q(x)}\right)^{1-p(x)} \leq 1+\frac{1-p(x)}{q(x)} \tag{2.3}$$

with equality iff $p(x) = q(x)$, $x \in \mathcal{X}$ .

If we multiply (2.3) by $q(x) \geq 0$ we get

$$\left(\frac{q\left(x\right)}{p(x)}\right)^{p(x)} \leq 1+q(x)-p(x), x \in \mathcal{X}. \tag{2.4}$$

Summing in (2.4) over $x \in \mathcal{X}$, we deduce

$$\sum_{x\in\mathcal{X}}\left(\frac{q\left(x\right)}{p(x)}\right)^{p(x)} \leq |\mathcal{X}|-\sum_{x\in\mathcal{X}}q(x)-\sum_{x\in\mathcal{X}}p(x)=|\mathcal{X}|$$

from where we deduce

$$\sum_{x\in\mathcal{X}}\left[1-\left(\frac{q\left(x\right)}{p(x)}\right)^{p(x)}\right] \geq 0$$

with equality iff $p(x) = q(x)$, $x \in \mathcal{X}$ .

Also, we have

$$\left(\frac{p(x)}{q(x)}\right)^{p(x)} = p(x)\left(\frac{1}{q(x)}\right)^{p(x)}\left(\frac{1}{p(x)}\right)^{1-p(x)}, x \in \mathcal{X}.$$

Using again the *weighted A-G-inequality* we can state that

$$\left(\frac{1}{q(x)}\right)^{p(x)}\left(\frac{1}{p(x)}\right)^{1-p(x)} \leq \frac{p(x)}{q(x)}+\frac{1}{p(x)}-1.$$

Multiplying this inequality by $p(x) \geq 0$ we get

$$\left(\frac{p(x)}{q(x)}\right)^{p(x)} \leq 1+\frac{p(x)^2}{q(x)}-p(x), x \in \mathcal{X}$$

which gives, by summation over $x$, that

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

i.e.,

$$\sum_{x\in\mathcal{X}}\left[\left(\frac{p(x)}{q(x)}\right)^{p(x)}-1\right] \leq \sum_{x\in\mathcal{X}}\frac{p(x)^2}{q(x)}-1$$

with equality iff $p(x) = q(x)$, $x \in \mathcal{X}$ .

As $\log_b(x) = \frac{\ln x}{\ln b}$, we deduce the desired inequality (2.1).

The following corollaries are important

**Corollary 7.** *Let $X$ be a random variable with the probability distribution $p(x)$, $x \in \mathcal{X}$ . Then we have the inequality*

$$\frac{1}{\ln b}\left[|\mathcal{X}|\sum_{x\in\mathcal{X}}p^2(x)-1\right] \geq \frac{1}{\ln b}\sum_{x\in\mathcal{X}}\left[|\mathcal{X}|^{p(x)}p(x)^{p(x)}-1\right] \tag{2.5}$$

$$\geq \log|\mathcal{X}| - H(X) \geq \frac{1}{\ln b}\sum_{x\in\mathcal{X}}\left[1 - \frac{1}{|\mathcal{X}|^{p(x)}\, p(x)^{p(x)}}\right] \geq 0.$$

*The equality holds in all inequalities iff* $p$ *is uniformly distributed on* $\mathcal{X}$ .

The proof follows by the above theorem choosing $q = u$, the uniform distribution on $\mathcal{X}$ .

The following corollary for mutual information also holds.

**Corollary 8.** *Let* $X$ *and* $Y$ *be as in the above theorem. Then*

$$\frac{1}{\ln b}\left[\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\frac{p^2(x,y)}{p(x)q(y)} - 1\right] \geq \frac{1}{\ln b}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\left[\left(\frac{p(x,y)}{p(x)q(y)}\right)^{p(x,y)} - 1\right] \tag{2.6}$$

$$\geq I(X,Y) \geq \frac{1}{\ln b}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\left[1 - \left(\frac{p(x,y)}{p(x)q(y)}\right)^{p(x,y)}\right] \geq 0.$$

*The equality holds in all inequalities iff* $X$ *and* $Y$ *are independent.*

For recent results in the applications of Theory of Inequalities in Information Theory and Coding, see the papers listed bellow.

## References


[1] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons Inc., 1991.

[2] S.S. Dragomir and N.M. Dragomir, An inequality for logarithms and its applications in coding theory, *Indian Journal of Mathematics* (in press).

[3] S. S. Dragomir and C. J. Goh, Some bounds on entropy measures in information theory, *Appl. Math. Lett.,* **10**(3) (1997) 23–28.

[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**(2) (1996) 1–11.

[5] M. Matić, C. E. M. Pearce and J. Pečarić, Improvements of some bounds on entropy measures in information theory, *Math. Inequal. and Appl.*, 1(1998) 295–304.

[6] M. Matić, C. E. M. Pearce and J. Pečarić, Refinements of some bounds in information theory, *J. Austr. Math. Soc. Ser. B,* to appear.

[7] S.S. Dragomir, N.M. Dragomir and K. Pranesh, Some Estimations of Kraft number and related results, *Soochow Journal of Mathematics*, **24**(4), 291-296.

[8] S.S. Dragomir, M. Sholz and J. Sunde, Some upper bounds for relative entropy and applications, *submitted.*

[9] S.S. Dragomir, Some remarks on the noiseless coding theorem, *submitted.*

[10] C. Calude and C. Grozea, Kraft-Chaitin inequality revisited, *J. UCS*, **5**(1996), 306-310.



School of Communications and Informatics, Victoria University of Technology
*E-mail address*: `sever@matilda.vu.edu.au`
*URL*: `http://matilda.vut.edu.au/~rgmia/`