An Inequality for Logarithmic Mapping and Applications for the Relative Entropy

Dragomir, Sever S (2000) An Inequality for Logarithmic Mapping and Applications for the Relative Entropy. RGMIA research report collection, 3 (2).

AN INEQUALITY FOR LOGARITHMIC MAPPING AND APPLICATIONS FOR THE RELATIVE ENTROPY

S.S. DRAGOMIR

Abstract. Using the concavity property of the log mapping and the weighted arithmetic mean - geometric mean inequality, we point out an analytic inequality for the logarithmic map and apply it for the Kullback-Leibler distance in Information Theory. Some applications for Shannon's entropy are given as well.

1. Introduction

Let $p(x), q(x), x \in \mathfrak{X}, card(\mathfrak{X}) < \infty$, be two probability mass functions. Define the *Kullback-Leibler distance* (see [1] or [2]) by

$$KL(p,q) := \sum_{x\in\mathfrak{X}} p(x) \log \frac{p(x)}{q(x)}, \tag{1.1}$$

the χ^2-*distance* (see for example [3]) by

$$D_{\chi^2}(p,q) := \sum_{x\in\mathfrak{X}} \frac{p^2(x) - q^2(x)}{q(x)} \tag{1.2}$$

and the *variation distance* (see for example [3]) by

$$V(p,q) := \sum_{x\in\mathfrak{X}} |p(x) - q(x)|. \tag{1.3}$$

The following theorem is of fundamental importance in Information Theory [4, p. 26].

Theorem 1. *(Information Inequality). Under the above assumptions for p and q, we have*

$$KL(p,q) \geq 0, \tag{1.4}$$

with equality iff $p(x) = q(x)$ for all $x \in \mathfrak{X}$.

As a matter of fact, the inequality (1.4) can be improved as follows (see [4, p. 300]):

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

$$KL(p,q) \geq \frac{1}{2} V^2(p,q) \geq 0, \tag{1.5}$$

with equality iff $p(x) = q(x)$ for all $x \in \mathfrak{X}$.

Date: 12th August, 1999.

1991 *Mathematics Subject Classification.* 26D15, 94Xxx.

Key words and phrases. Analytic Inequalities, Kull-back-Leibler distances, χ^2-distance, Variation distance.

In [5] (see also [6]), the authors proved the following counterpart of (1.5).

Theorem 3. *Let $p(x), q(x) > 0, x \in \mathfrak{X}$ be two probability mass functions. Then*

$$D_{\chi^2}(p,q) \geq KL(p,q) \geq 0, \tag{1.6}$$

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

In the same paper [6], the authors applied (1.6) for Shannon's entropy, mutual information, etc....

In the present paper, we point out an improvement of (1.6) and apply it in the same manner as in [6].

2. An Elementary Inequality

The following analytic inequality for the logarithmic function holds.

Theorem 4. *Let $a \in (0,1)$ and $b \in (0,\infty)$. Then we have the inequality*

$$\frac{a^2}{b} - a \geq \left(\frac{a}{b}\right)^a - 1 \geq a\ln a - a\ln b \geq 1 - \left(\frac{b}{a}\right)^a \geq a - b. \tag{2.1}$$

The equality holds in each inequality iff $a = b$.

Proof. We know that for a differentiable strictly convex mapping $f : I \to \mathbb{R}$, we have the double inequality

$$f'(x)(x-y) \geq f(x) - f(y) \geq f'(y)(x-y) \tag{2.2}$$

for all $x, y \in I$, $x \leq y$. The equality holds in (2.2) iff $x = y$.

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

$$\frac{1}{y}(x-y) \geq \ln x - \ln y \geq \frac{1}{x}(x-y) \tag{2.3}$$

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

Choose in (2.3) $x = a^a$ and $y = b^a$ to get

$$\left(\frac{a}{b}\right)^a - 1 \geq a\ln a - a\ln b \geq 1 - \left(\frac{b}{a}\right)^a; \ a, b > 0,$$

with equality iff $a = b$, and the second and third inequalities in (2.1) are proved.

Further, we are going to use the weighted *arithmetic mean - geometric mean inequality* for two positive numbers, i.e., we recall

$$\alpha^t \beta^{1-t} \leq t\alpha + (1-t)\beta \text{ for } \alpha, \beta > 0 \text{ and } t \in (0,1), \tag{2.4}$$

with equality iff $\alpha = \beta$.

Choose $\alpha = \frac{1}{a}$, $\beta = \frac{1}{b}$ and $t = a$ in (2.4) to obtain

$$\left(\frac{1}{a}\right)^a \left(\frac{1}{b}\right)^{1-a} \leq a \cdot \frac{1}{a} + (1-a) \cdot \frac{1}{b}$$

with equality iff $a = b$, which is equivalent to

$$\left(\frac{1}{a}\right)^a \left(\frac{1}{b}\right)^{1-a} \leq 1 + \frac{1-a}{b}. \tag{2.5}$$

If we multiply (2.5) by $b>0$, we have

$$\left(\frac{b}{a}\right)^a \le 1+b-a,$$

with equality iff $a=b$, and the last inequality in (2.1) is proved.

In addition, if we choose in (2.4) $\alpha=\frac{1}{b}$, $\beta=\frac{1}{a}$ and $t=a$, we obtain

$$\left(\frac{1}{b}\right)^a\left(\frac{1}{a}\right)^{1-a} \le \frac{a}{b}+\frac{1}{a}-1, \tag{2.6}$$

with equality iff $a=b$.

If we multiply (2.6) by $a>0$, then we get

$$\frac{a^a}{b^a} \le \frac{a^2}{b}-a+1$$

with equality iff $a=b$, which is the first inequality in (2.1). ∎

3. Inequalities for Sequences of Positive Numbers

The following inequality for sequences of positive numbers holds.

Theorem 5. *Let $a_i\in(0,1)$ and $b_i>0$ $(i=1,...,n)$. If $p_i>0$ $(i=1,...,n)$ is such that $\sum_{i=1}^n p_i=1$, then we have*

$$\begin{aligned} &\exp\left[\sum_{i=1}^n p_i\frac{a_i^2}{b_i}-\sum_{i=1}^n p_ia_i\right] \\ \ge\ &\exp\left[\sum_{i=1}^n p_i\left(\frac{a_i}{b_i}\right)^{a_i}-1\right] \ge \prod_{i=1}^n\left(\frac{a_i}{b_i}\right)^{a_ip_i} \\ \ge\ &\exp\left[1-\sum_{i=1}^n p_i\left(\frac{b_i}{a_i}\right)^{a_i}\right] \ge \exp\left[\sum_{i=1}^n p_ia_i-\sum_{i=1}^n p_ib_i\right], \end{aligned} \tag{3.1}$$

with equality iff $a_i=b_i$ for all $i\in\{1,...,n\}$.

Proof. Choose in (3.1) $a=a_i$, $b=b_i$ $(i=1,...,n)$ to obtain

$$\frac{a_i^2}{b_i}-a_i \ge \left(\frac{a_i}{b_i}\right)^{a_i}-1 \ge a_i\ln a_i-a_i\ln b_i \ge 1-\left(\frac{b_i}{a_i}\right)^{a_i} \ge a_i-b_i \tag{3.2}$$

for all $i\in\{1,...,n\}$.

Multiplying (3.2) by $p_i>0$ and summing over i from 1 up to n, we get

$$\begin{aligned} &\sum_{i=1}^n p_i\frac{a_i^2}{b_i}-\sum_{i=1}^n p_ia_i \\ \ge\ &\sum_{i=1}^n p_i\left(\frac{a_i}{b_i}\right)^{a_i}-1 \ge \sum_{i=1}^n p_ia_i\ln\left(\frac{a_i}{b_i}\right) \\ \ge\ &1-\sum_{i=1}^n p_i\left(\frac{b_i}{a_i}\right)^{a_i} \ge \sum_{i=1}^n p_ia_i-\sum_{i=1}^n p_ib_i, \end{aligned} \tag{3.3}$$

which is equivalent to (3.1).

The case of equality follows from the fact that in each of the inequalities (3.2), we have an equality iff $a_i=b_i$ for all $i\in\{1,...,n\}$. ∎

The following corollary is obvious.

Corollary 1. *With the above assumptions for a_i, b_i $(i=1,...,n)$, we have the inequality*

$$\begin{aligned}(3.4)\qquad & \exp\left(\frac{1}{n}\sum_{i=1}^{n}\frac{a_i^2}{b_i}-\frac{1}{n}\sum_{i=1}^{n}a_i\right)\\ \geq\; & \exp\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{a_i}{b_i}\right)^{a_i}-1\right]\geq\sqrt[n]{\prod_{i=1}^{n}\left(\frac{a_i}{b_i}\right)^{a_i}}\\ \geq\; & \exp\left[1-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{b_i}{a_i}\right)^{a_i}\right]\geq\exp\left(\frac{1}{n}\sum_{i=1}^{n}a_i-\frac{1}{n}\sum_{i=1}^{n}b_i\right),\end{aligned}$$

with equality iff $a_i=b_i$ for all $i\in\{1,...,n\}$.

Another result for sequences of positive numbers is the following one.

Theorem 6. *Let $a_i\in(0,1)$ $(i=1,...,n)$ and $b_j>0$ $(j=1,...,m)$. If $p_i>0$ $(i=1,...,n)$ is such that $\sum_{i=1}^{n}p_i=1$ and $q_j>0$ $(j=1,...,m)$ is such that $\sum_{j=1}^{m}q_j=1$, then we have the inequality*

$$\begin{aligned}(3.5)\qquad & \exp\left(\sum_{i=1}^{n}p_ia_i^2\sum_{j=1}^{m}\frac{q_j}{b_j}-\sum_{i=1}^{n}p_ia_i\right)\\ \geq\; & \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{m}p_iq_j\left(\frac{a_i}{b_j}\right)^{a_i}-1\right]\geq\frac{\prod_{i=1}^{n}a_i^{a_ip_i}}{\prod_{j=1}^{m}\left(b_j^{q_j}\right)^{\sum_{i=1}^{n}p_ia_i}}\\ \geq\; & \exp\left[1-\sum_{i=1}^{n}\sum_{j=1}^{m}p_iq_j\left(\frac{b_i}{a_i}\right)^{a_i}\right]\geq\exp\left(\sum_{i=1}^{n}p_ia_i-\sum_{j=1}^{m}q_jb_j\right).\end{aligned}$$

The equality holds in (3.5) iff $a_1=...=a_n=b_1=...=b_m$.

Proof. Using the inequality (2.1), we can state that

$$(3.6)\qquad \frac{a_i^2}{b_j}-a_i\geq\left(\frac{a_i}{b_j}\right)^{a_i}-1\geq a_i\ln a_i-a_i\ln b_j\geq 1-\left(\frac{b_j}{a_i}\right)^{a_i}\geq a_i-b_j$$

for all $i\in\{1,...,n\}$ and $j\in\{1,...,m\}$.

Multiplying (3.6) by $p_iq_j>0$ and summing over i from 1 to n and over j from 1 to m, we deduce

$$\begin{aligned} & \sum_{i=1}^{n}p_ia_i^2\sum_{j=1}^{m}\frac{q_j}{b_j}-\sum_{i=1}^{n}p_ia_i\\ \geq\; & \sum_{i=1}^{n}\sum_{j=1}^{m}p_iq_j\left(\frac{a_i}{b_j}\right)^{a_i}-1\geq\sum_{i=1}^{n}p_ia_i\ln a_i-\sum_{i=1}^{n}p_ia_i\sum_{j=1}^{m}q_j\ln b_j\\ \geq\; & 1-\sum_{i=1}^{n}\sum_{j=1}^{m}p_iq_j\left(\frac{b_i}{a_i}\right)^{a_i}\geq\sum_{i=1}^{n}p_ia_i-\sum_{j=1}^{m}q_jb_j,\end{aligned}$$

which is clearly equivalent to (3.5).

The case of equality follows from the fact that in each of inequalities in (3.6), we have an equality iff $a_i = b_j$ for all $i \in \{1,...,n\}$ and $j \in \{1,...,m\}$, which is equivalent to $a_1 = ... = a_n = b_1 = ... = b_m$. ∎

The following corollary holds.

Corollary 2. *Under the above assumptions for a_i, b_j, we have the inequality*

$$\begin{aligned} (3.7) \qquad & \exp\left(\frac{1}{n}\sum_{i=1}^{n} a_i^2 \frac{1}{m}\sum_{j=1}^{m}\frac{1}{b_j} - \frac{1}{n}\sum_{i=1}^{n} a_i\right) \\ \geq\; & \exp\left[\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{a_i}{b_j}\right)^{a_i} - 1\right] \geq \frac{\sqrt[n]{\prod_{i=1}^{n} a_i^{a_i}}}{\sqrt[nm]{\prod_{j=1}^{m} b_j^{\sum_{i=1}^{n} a_i}}} \\ \geq\; & \exp\left[1 - \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{b_j}{a_i}\right)^{a_i}\right] \geq \exp\left(\frac{1}{n}\sum_{i=1}^{n} a_i - \frac{1}{m}\sum_{j=1}^{m} b_j\right), \end{aligned}$$

with equality iff $a_1 = ... = a_n = b_1 = ... = b_m$.

4. Some Inequalities for Distance Functions

In 1951, Kullback and Leibler introduced the following *distance function* in Information Theory (see [2] or [3])

$$KL(p,q) := \sum_{i=1}^{n} p_i \log\frac{p_i}{q_i}, \tag{4.1}$$

provided that $p,q \in \mathbb{R}_{++}^n := \{x = (x_1,...,x_n) \in \mathbb{R}^n,\ x_i > 0,\ i = 1,...,n\}$.

Another useful distance function is the χ^2-*distance* given by (see [3])

$$D_{\chi^2}(p,q) := \sum_{i=1}^{n} \frac{p_i^2 - q_i^2}{q_i}, \tag{4.2}$$

where $p,q \in \mathbb{R}_{++}^n$.

In this section, we introduce the following two new distance functions

$$P_2(p,q) := \sum_{i=1}^{n}\left[\left(\frac{p_i}{q_i}\right)^{p_i} - 1\right] \tag{4.3}$$

and

$$P_1(p,q) := \sum_{i=1}^{n}\left[\left(\frac{q_i}{p_i}\right)^{p_i} - 1\right], \tag{4.4}$$

provided $p,q \in \mathbb{R}_{++}^n$.

The following inequality connecting all the above four distance functions holds.

Theorem 7. *Let $p,q \in \mathbb{R}_{++}^n$ with $p_i \in (0,1)$. Then we have the inequality:*

$$D_{\chi^2}(p,q) + Q_n - P_n \geq P_2(p,q) \geq KL(p,q) \geq P_1(p,q) \geq P_n - Q_n, \tag{4.5}$$

where $P_n := \sum_{i=1}^{n} p_i = 1$, $Q_n := \sum_{i=1}^{n} q_i$.
The equality holds in (4.5) iff $p_i = q_i$ for all $i \in \{1,...,n\}$.

Proof. Apply inequality (2.1) for $a = p_i$, $b = q_i$ to get

$$\frac{p_i^2}{q_i} - p_i \geq \left(\frac{p_i}{q_i}\right) \geq p_i \ln p_i - p_i \ln q_i \geq 1 - \left(\frac{q_i}{p_i}\right)^{p_i} \geq p_i - q_i \tag{4.6}$$

for all $i \in \{1, ..., n\}$.

Summing over i from 1 to n, we have

$$\sum_{i=1}^{n} \frac{p_i^2}{q_i} - P_n \geq P_2(p,q) \geq KL(p,q) \geq P_1(p,q) \geq P_n - Q_n.$$

However, it is easy to see that

$$\sum_{i=1}^{n} \frac{p_i^2}{q_i} - Q_n + Q_n - P_n = D_{\chi^2}(p,q) + Q_n - P_n$$

and the inequality (4.5) is obtained.

The case of equality is also obvious by Theorem 4. ∎

Corollary 3. *Let p, q be a probability distribution. Then we have the inequality:*

$$D_{\chi^2}(p,q) \geq P_2(p,q) \geq KL(p,q) \geq P_1(p,q) \geq 0. \tag{4.7}$$

The equality holds in (4.7) iff $p = q$.

The proof is obvious by Theorem 7, on observing that for p, q as probability distributions we have $P_n = Q_n = 1$.

5. Applications for Shannon's Entropy

The *entropy* of a random variable is a measure of the uncertainty of the random variable, it is a measure of the amount of information required on the average to describe the random variable.

Let $p(x)$, $x \in \mathcal{X}$ be a probability mass function. Define the *Shannon's entropy* f of a random variable X having the probability distribution p by

$$H(X) := \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}. \tag{5.1}$$

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

Now assume that $|\mathcal{X}|$ ($card(\mathcal{X}) = |\mathcal{X}|$) is finite and let $u(x) = \frac{1}{|\mathcal{X}|}$ be the uniform probability mass function in $\mathcal{X}$. It is well known that [4, p. 27]

$$\begin{aligned} KL(p,q) &= \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) \\ &= \log|\mathcal{X}| - H(X). \end{aligned} \tag{5.2}$$

The following result is important in Information Theory [4, p. 27]

Theorem 8. *Let X, p and $\mathcal{X}$ be as above. Then*

$$H(X) \leq \log|\mathcal{X}|, \tag{5.3}$$

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

In what follows, by the use of Corollary 3, we are able to point out the following estimate for the difference $\log|\mathcal{X}| - H(X)$.

Theorem 9. *Let X, p and $\mathcal{X}$ be as above. Then*

$$\begin{aligned} |\mathcal{X}|\, E(X) - 1 &\geq \sum_{x\in\mathcal{X}} \left[|\mathcal{X}|^{p(x)} [p(x)]^{p(x)} - 1 \right] \\ &\geq \ln |\mathcal{X}| - H(X) \\ &\geq \sum_{x\in\mathcal{X}} \left[|\mathcal{X}|^{-p(x)} [p(x)]^{-p(x)} - 1 \right] \geq 0, \end{aligned} \tag{5.4}$$

where $E(X)$ is the informational energy of X, i.e., $E(X) := \sum_{x\in\mathcal{X}} p^2(x)$. The equality holds in (5.4) iff $p(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$.

The proof is obvious by Corollary 3 by choosing $u(x) = \frac{1}{|\mathcal{X}|}$.

6. Applications for Mutual Information

We consider *mutual information*, which is a measure of the amount of information that one random variable contains about another random variable. It is the reduction of uncertainty of one random variable due to the knowledge of the other [4, p. 18].

To be more precise, consider two random variables X and Y with a joint probability mass function $r(x, y)$ and marginal probability mass functions $p(x)$ and $q(y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. The mutual information is the relative entropy between the joint distribution and the product distribution, that is,

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

The following result is well known [4, p. 27].

Theorem 10. *(Non-negativity of mutual information) For any two random variables X, Y*

$$I(X,Y) \geq 0, \tag{6.1}$$

with equality iff X and Y are independent.

In what follows, by the use of Corollary 3, we are able to point out the following estimate for the mutual information.

Theorem 11. *Let X and Y be as above. Then we have the inequality*

$$\begin{aligned} &\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{r^2(x,y)}{p(x)\, q(y)} - 1 \\ \geq\ &\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \left[\left(\frac{r(x,y)}{p(x)\, q(y)} \right)^{r(x,y)} - 1 \right] \geq I(X,Y) \\ \geq\ &\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \left[1 - \left(\frac{r(x,y)}{p(x)\, q(y)} \right)^{r(x,y)} \right] \geq 0. \end{aligned}$$

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

References

[1] S. KULLBACK and R.A. LEIBLER, On information and sufficiency, *Annals Maths. Statist.,* **22** (1951), 79 - 86.

[2] S. KULLBACK, *Information Theory and Statistics,* J. Wiley, New York, 1959.

[3] A. BEN-TAL, A. BEN-ISRAEL and M. TEBOULLE, Certainty equivalents and information measures: duality and extremal, *J. Math. Anal. Appl.,* **157** (1991), 211 - 236.

[4] T.M. COVER and J.A. THOMAS, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[5] S.S. DRAGOMIR and N.M. DRAGOMIR, An inequality for logarithms and its application in coding theory, *Indian J. of Math.* (in press).

[6] S.S. DRAGOMIR, M. SCHOLZ and J. SUNDE, Some upper bounds for relative entropy and applications, *Computers and Mathematics with Applications.*(in press).

[7] D.S. MITRINOVIĆ, *Analytic Inequalities,* Springer Verlag, 1970.

[8] M. MATIĆ, *Jensen's Inequality and Applications in Information Theory (* in Croatian), Ph.D Dissertation, Split, Croatia.

School of Communications and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City MC 80001, Victoria, Australia

E-mail address: `sever@matilda.vu.edu.au`

URL: `http://melba.vu.edu.au/~rgmia`