Journal of Inequalities in Pure and Applied Mathematics

http://jipam.vu.edu.au/

Volume 2, Issue 1, Article 11, 2001

ON SOME APPLICATIONS OF THE AG INEQUALITY IN INFORMATION THEORY

B. MOND AND J. PEČARIĆ

SCHOOL OF MATHEMATICS, LA TROBE UNIVERSITY, BUNDOORA 3086, VICTORIA, AUSTRALIA.
b.mond@latrobe.edu.au

FACULTY OF TEXTILE TECHNOLOGY, UNIVERSITY OF ZAGREB, ZAGREB, CROATIA
pecaric@hazu.hr

Received 3 November, 2000; accepted 11 January, 2001
Communicated by A. Lupas

ABSTRACT. Recently, S.S. Dragomir used the concavity property of the log mapping and the weighted arithmetic mean-geometric mean inequality to develop new inequalities that were then applied to Information Theory. Here we extend these inequalities and their applications.

Key words and phrases: Arithmetic-Geometric Mean, Kullback-Leibler Distances, Shannon's Entropy.

2000 *Mathematics Subject Classification.* 26D15.

1. INTRODUCTION

One of the most important inequalities is the arithmetic-geometric means inequality:

Let $a_i,\ p_i,\ i = 1, \dots, n$ be positive numbers, $P_n = \sum_{i=1}^n p_i$. Then

$$\prod_{i=1}^{n} a_i^{p_i/P_n} \leq \frac{1}{P_n} \sum_{i=1}^{n} p_i a_i, \tag{1.1}$$

with equality iff $a_1 = \dots = a_n$.

It is well-known that using (1.1) we can prove the following generalization of another well-known inequality, that is Hölder's inequality:

Let $p_{ij},\ q_i\ (i = 1, \dots, m;\ j = 1, \dots, n)$ be positive numbers with $Q_m = \sum_{i=1}^m q_i$. Then

$$\sum_{j=1}^{n} \prod_{i=1}^{m} (p_{ij})^{\frac{q_i}{Q_m}} \leq \prod_{i=1}^{m} \left(\sum_{j=1}^{n} p_{ij} \right)^{\frac{q_i}{Q_m}}. \tag{1.2}$$

In this note, we show that using (1.1) we can improve some recent results which have applications in information theory.

ISSN (electronic): 1443-5756

042-00

2. AN INEQUALITY OF I.A. ABOU-TAIR AND W.T. SULAIMAN

The main result from [1] is:

Let $p_{ij},\ q_i\ (i=1,\dots,m;\ j=1,\dots,n)$ be positive numbers. Then

$$\sum_{j=1}^{n}\prod_{i=1}^{m}(p_{ij})^{\frac{q_i}{Q_m}} \le \frac{1}{Q_m}\sum_{i=1}^{m}\sum_{j=1}^{n}p_{ij}q_i\,. \tag{2.1}$$

Moreover, set in (1.1), $n=m,\ p_i=q_i,\ a_i=\sum_{j=1}^{n}p_{ij}$. We now have

$$\prod_{i=1}^{m}\left(\sum_{j=1}^{n}p_{ij}\right)^{\frac{q_i}{Q_m}} \le \frac{1}{Q_m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}p_{ij}q_i\right). \tag{2.2}$$

Now (1.2) and (2.2) give

$$\sum_{j=1}^{n}\prod_{i=1}^{m}(p_{ij})^{\frac{q_i}{Q_m}} \le \prod_{i=1}^{m}\left(\sum_{j=1}^{n}p_{ij}\right)^{\frac{q_i}{Q_m}} \le \frac{1}{Q_m}\sum_{i=1}^{m}\sum_{j=1}^{n}p_{ij}q_i. \tag{2.3}$$

which is an interpolation of (2.1). Moreover, the generalized Hölder inequality was obtained in [1] as a consequence of (2.1). This is not surprising since (2.1), for $n=1$, becomes

$$\prod_{i=1}^{m}(p_{i1})^{\frac{q_i}{Q_m}} \le \frac{1}{Q_m}\sum_{i=1}^{m}p_{i1}q_i$$

which is, in fact, the A-G inequality (1.1) (set $m=n,\ p_{i1}=a_i$ and $q_i=p_i$). Theorem 3.1 in [1] is the well-known Shannon inequality:

Given $\sum_{i=1}^{n}a_i=a,\ \sum_{i=1}^{n}b_i=b$. Then

$$a\ln\left(\frac{a}{b}\right) \le \sum_{i=1}^{n}a_i\ln\left(\frac{a_i}{b_i}\right);\quad a_i,\ b_i>0.$$

It was obtained from (2.1) through the special case

$$\prod_{i=1}^{n}\left(\frac{b_i}{a_i}\right)^{\frac{a_i}{a}} \le \frac{b}{a}. \tag{2.4}$$

Let us note that (2.4) is again a direct consequence of the A-G inequality. Indeed, in (1.1), setting $a_i\to b_i/a_i,\ p_i\to a_i,\ i=1,\dots,n$ we have (2.4). Theorem 3.2 from [1] is Rényi's inequality. Given $\sum_{i=1}^{m}a_i=a,\ \sum_{i=1}^{m}b_i=b$, then for $\alpha>0,\alpha\ne 1$,

$$\frac{1}{\alpha-1}(a^{\alpha}b^{1-\alpha}-a) \le \sum_{i=1}^{m}\frac{1}{\alpha-1}\left(a_i^{\alpha}b_i^{1-\alpha}-a_i\right);\ a_i,b_i\ge 0.$$

In fact, in the proof given in [1], it was proved that Hölder's inequality is a consequence of (2.1). As we have noted, Hölder's inequality is also a consequence of the A-G inequality.

3. ON SOME INEQUALITIES OF S.S. DRAGOMIR

The following theorems were proved in [2]:

Theorem 3.1. *Let $a_i \in (0,1)$ and $b_i > 0 (i = 1, \dots, n)$. If $p_i > 0$ $(i = 1, \dots, n)$ is such that $\sum_{i=1}^n p_i = 1$, then*

$$\exp\left[\sum_{i=1}^n p_i \frac{a_i^2}{b_i} - \sum_{i=1}^n p_i a_i\right] \geq \exp\left[\sum_{i=1}^n p_i \left(\frac{a_i}{b_i}\right)^{a_i} - 1\right] \tag{3.1}$$
$$\geq \prod_{i=1}^n \left(\frac{a_i}{b_i}\right)^{a_i p_i}$$
$$\geq \exp\left[1 - \sum_{i=1}^n p_i \left(\frac{p_i}{a_i}\right)^{a_i}\right]$$
$$\geq \exp\left[\sum_{i=1}^n p_i a_i - \sum_{i=1}^n p_i b_i\right]$$

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

Theorem 3.2. *Let $a_i \in (0,1)$ $(i = 1, \dots, n)$ and $b_j > 0$ $(j = 1, \dots, m)$. If $p_i > 0$ $(i = 1, \dots, n)$ is such that $\sum_{i=1}^n p_i = 1$ and $q_j > 0$ $(j = 1, \dots, m)$ is such that $\sum_{j=1}^m q_j = 1$, then we have the inequality*

$$\exp\left(\sum_{i=1}^n p_i a_i^2 \sum_{j=1}^m \frac{q_j}{b_j} - \sum_{i=1}^n p_i a_i\right) \geq \exp\left[\sum_{i=1}^n \sum_{j=1}^m p_i q_j \left(\frac{a_i}{b_j}\right)^{a_i} - 1\right] \tag{3.2}$$
$$\geq \frac{\prod_{i=1}^n a_i^{a_i p_i}}{\prod_{j=1}^m \left(b_j^{q_j}\right)^{\sum_{i=1}^n p_i a_i}}$$
$$\geq \exp\left[1 - \sum_{i=1}^n \sum_{j=1}^m p_i q_j \left(\frac{b_j}{a_i}\right)^{a_i}\right]$$
$$\geq \exp\left(\sum_{i=1}^n p_i a_i - \sum_{j=1}^m q_j b_j\right).$$

The equality holds in (3.2) iff $a_1 = \cdots = a_n = b_1 = \cdots = b_m$.

First we give an improvement of the second and third inequality in (3.1).

Theorem 3.3. *Let a_i, b_i and p_i $(i = 1, \dots, n)$ be positive real numbers with $\sum_{i=1}^n p_i = 1$. Then*

$$\exp\left[p_i \left(\frac{a_i}{b_i}\right)^{a_i} - 1\right]^{-1} \geq \sum_{i=1}^n p_i \left(\frac{a_i}{b_i}\right)^{a_i} \tag{3.3}$$
$$\geq \prod_{i=1}^n \left(\frac{a_i}{b_i}\right)^{p_i a_i}$$
$$\geq \left[\sum_{i=1}^n p_i \left(\frac{b_i}{a_i}\right)^{a_i}\right]^{-1}$$
$$\geq \exp\left[1 - \sum_{i=1}^n p_i \left(\frac{b_i}{a_i}\right)^{a_i}\right],$$

with equality iff $a_i = b_i,\ i = 1, \dots, n$.

Proof. The first inequality in (3.3) is a simple consequence of the following well-known elementary inequality

$$e^{x-1} \geq x, \quad \text{for all } x \in \mathbb{R} \tag{3.4}$$

with equality iff $x = 1$. The second inequality is a simple consequence of the A-G inequality that is, in (1.1), set $a_i \to (a_i/b_i)^{a_i},\ i = 1, \ldots, n$. The third inequality is again a consequence of (1.1). Namely, for $a_i \to (b_i/a_i)^{a_i},\ i = 1, \ldots, n$, (1.1) becomes

$$\prod_{i=1}^{n} \left(\frac{b_i}{a_i}\right)^{a_i p_i} \leq \sum_{i=1}^{n} p_i \left(\frac{b_i}{a_i}\right)^{a_i}$$

which is equivalent to the third inequality. The last inequality is again a consequence of (3.4). ∎

Theorem 3.4. *Let $a_i \in (0,1)$ and $b_i > 0$ $(i = 1, \ldots, n)$. If $p_i > 0,\ i = 1, \ldots, n$ is such that $\sum_{i=1}^{n} p_i = 1$, then*

$$\begin{aligned} \exp\left[\sum_{i=1}^{n} p_i \left(\frac{a_i^2}{b_i}\right) - \sum_{i=1}^{n} p_i a_i\right] &\geq \exp\left[\sum_{i=1}^{n} p_i \left(\frac{a_i}{b_i}\right)^{a_i} - 1\right] \\ &\geq \sum_{i=1}^{n} p_i \left(\frac{a_i}{b_i}\right)^{a_i} \\ &\geq \prod_{i=1}^{n} \left(\frac{a_i}{b_i}\right)^{p_i a_i} \\ &\geq \left[\sum_{i=1}^{n} p_i \left(\frac{b_i}{a_i}\right)^{a_i}\right]^{-1} \\ &\geq \exp\left[1 - \sum_{i=1}^{n} p_i \left(\frac{b_i}{a_i}\right)^{a_i}\right] \\ &\geq \exp\left[\sum_{i=1}^{n} p_i a_i - \sum_{i=1}^{n} p_i b_i\right] \end{aligned} \tag{3.5}$$

with equality iff $a_i = b_i$ for all $i = 1, \ldots, n$.

Proof. The theorem follows from Theorems 3.1 and 3.3. ∎

Theorem 3.5. *Let $a_i,\ p_i$ $(i = 1, \ldots, n)$; $b_j,\ q_j$ $(j = 1, \ldots, m)$ be positive numbers with $\sum_{i=1}^{n} p_i = \sum_{j=1}^{m} q_j = 1$. Then*

$$\begin{aligned} \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{a_i}{b_j}\right)^{a_i} - 1\right] &\geq \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{a_i}{b_j}\right)^{a_i} \\ &\geq \frac{\prod_{i=1}^{n} a_i^{a_i p_i}}{\prod_{j=1}^{m} \left(b_j^{q_j}\right)^{\sum_{i=1}^{n} p_i a_i}} \\ &\geq \left[\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{b_j}{a_i}\right)^{a_i}\right]^{-1} \end{aligned} \tag{3.6}$$

$$\geq \exp\left[1 - \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{b_j}{a_i}\right)^{a_i}\right]^{-1}.$$

Equality in (3.6) holds iff $a_1 = \cdots = a_n = b_1 = \cdots = b_m$.

Proof. The first and the last inequalities are simple consequences of (3.4). The second is also a simple consequence of the A-G inequality. Namely, we have

$$\frac{\prod_{i=1}^{n} a_i^{a_i p_i}}{\prod_{j=1}^{m} \left(b_j^{q_j}\right)^{\sum_{i=1}^{n} p_i a_i}} = \prod_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{a_i}{b_j}\right)^{a_i p_i q_j} \leq \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{a_i}{b_j}\right)^{a_i},$$

which is the second inequality in (3.6). By the A-G inequality, we have

$$\prod_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{b_j}{a_i}\right)^{a_i p_i q_j} \leq \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{b_j}{a_i}\right)^{a_i}$$

which gives the third inequality in (3.6). ∎

Theorem 3.6. *Let the assumptions of Theorem 3.2 be satisfied. Then*

$$\begin{aligned}
\exp\left[\sum_{i=1}^{n} p_i a_i^2 \sum_{j=1}^{m}\left(\frac{q_j}{b_j}\right) - \sum_{i=1}^{n} p_i a_i\right] &\geq \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{a_i}{b_j}\right)^{a_i} - 1\right] \\
&\geq \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{a_i}{b_j}\right)^{a_i} \\
&\geq \frac{\prod_{i=1}^{n} a_i^{a_i p_i}}{\prod_{j=1}^{m} \left(b_j^{q_j}\right)^{\sum_{i=1}^{n} p_i a_i}} \\
&\geq \left[\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{b_j}{a_i}\right)^{a_i}\right]^{-1} \\
&\geq \exp\left[1 - \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\frac{b_j}{a_i}\right)^{a_i}\right] \\
&\geq \exp\left[\sum_{i=1}^{n} p_i a_i - \sum_{j=1}^{m} q_j b_j\right].
\end{aligned} \tag{3.7}$$

Equality holds in (3.7) iff $a_1 = \cdots = a_n = b_1 = \cdots = b_m$.

Proof. The theorem is a simple consequence of Theorems 3.2 and 3.5. ∎

4. SOME INEQUALITIES FOR DISTANCE FUNCTIONS

In 1951, Kullback and Leibler introduced the following *distance function* in Information Theory (see [4] or [5])

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

provided that $p, q \in \mathbb{R}^n_{++} := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n,\ x_i > 0,\ i = 1, \dots, n\}$. Another useful distance function is the χ^2-*distance* given by (see [5])

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

where $p, q \in \mathbb{R}^n_{++}$. S.S. Dragomir [2] introduced 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 4.1. *Let $p, q \in \mathbb{R}^n_{++}$ with $p_i \in (0,1)$. Then we have the inequality:*

$$\begin{aligned} D_{\chi^2}(p,q) + Q_n - P_n &\geq P_2(p,q) \\ &\geq n \ln \left[\left(\frac{1}{n} \right) P_2(p,q) + 1 \right] \\ &\geq KL(p,q) \\ &\geq -n \ln \left[- \left(\frac{1}{n} \right) P_1(p,q) + 1 \right] \\ &\geq P_1(p,q) \\ &\geq P_n - Q_n, \end{aligned} \tag{4.5}$$

where $P_n = \sum_{i=1}^n p_i = 1,\ Q_n = \sum_{i=1}^n q_i$. Equality holds in (4.5) iff $p_i = q_i$ $(i = 1, \dots, n)$.

Proof. Set in (3.5), $p_i = 1/n,\ a_i = p_i,\ b_i = q_i$ $(i = 1, \dots, n)$ and take logarithms. After multiplication by n, we get (4.5). ∎

Corollary 4.2. *Let p, q be probability distributions. Then we have*

$$\begin{aligned} D_{\chi^2}(p,q) &\geq P_2(p,q) \\ &\geq n \ln \left[\left(\frac{1}{n} \right) P_2(p,q) + 1 \right] \\ &\geq KL(p,q) \\ &\geq -n \ln \left[1 - \left(\frac{1}{n} \right) P_1(p,q) \right] \\ &\geq P_1(p,q) \geq 0. \end{aligned} \tag{4.6}$$

Equality holds in (4.6) iff $p = q$.

Remark 4.3. Inequalities (4.5) and (4.6) are improvements of related results in [2].

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

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

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 $|\chi|$ $(card(\chi) = |\chi|)$ is finite and let $u(x) = \frac{1}{|\chi|}$ be the uniform probability mass function in χ. It is well known that [5, p. 27]

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

The following result is important in Information Theory [5, p. 27]:

Theorem 5.1. *Let X, p and χ be as above. Then*

$$H(X) \le \log|\chi|, \tag{5.3}$$

with equality if and only if X has a uniform distribution over χ.

In what follows, by the use of Corollary 4.2, we are able to point out the following estimate for the difference $\log|\chi| - H(X)$, that is, we shall give the following improvement of Theorem 9 from [2]:

Theorem 5.2. *Let $X,\ p$ and χ be as above. Then*

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

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

Proof. The proof is obvious by Corollary 4.2 by choosing $u(x) = \frac{1}{|\chi|}$. ∎

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 [6, 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,

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

The following result is well known [6, p. 27].

Theorem 6.1. *(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 4.2, we are able to point out the following estimate for the mutual information, that is, the following improvement of Theorem 11 of [2]:

Theorem 6.2. *Let X and Y be as above. Then we have the inequality*

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

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

REFERENCES

[1] I.A. ABOU-TAIR AND W.T. SULAIMAN, Inequalities via convex functions, *Internat. J. Math. and Math. Sci.*, **22** (1999), 543–546.

[2] S.S. DRAGOMIR, An inequality for logarithmic mapping and applications for the relative entropy, *RGMIA Res. Rep. Coll.*, **3**(2) (2000), Article 1, `http://rgmia.vu.edu.au/v3n2.html`

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

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

[5] 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.

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