An Inequality for Logarithms and its Application in Coding Theory

RGMIA Research Report Collection, Vol. 1, No. 1, 1998
http://sci.vut.edu.au/∼rgmia/reports.html

AN INEQUALITY FOR LOGARITHMS AND ITS APPLICATION IN CODING THEORY

N.M. DRAGOMIR AND S.S. DRAGOMIR

Abstract. In this paper we prove a new analytic inequality for logarithms and apply it for the Noiseless Coding Theorem.

1 Introduction

The following analytic inequality for logarithms is well known in the literature (see for example [1, Lemma 1.2.2, p. 22]):

Lemma 1.1. *Let $P = (p_1, ..., p_n)$ be a probability distribution, that is, $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$. Let $Q = (q_1, ..., q_n)$ have the property that $0 \leq q_i \leq 1$ and $\sum_{i=1}^{n} q_i \leq 1$ (note the inequality here). Then*

$$\sum_{i=1}^{n} p_i \log_b \left(\frac{1}{p_i}\right) \leq \sum_{i=1}^{n} p_i \log_b \left(\frac{1}{q_i}\right) \tag{1.1}$$

where $b > 1$, $0 \cdot \log_b(1/0) = 0$ and $p \cdot \log_b (1/0) = +\infty$. Furthermore, equality holds if and only if $q_i = p_i$ for all $i \in \{1, ..., n\}$.

Note that the proof of this fact uses the elementary inequality for logarithms (see [1, p. 22])

$$\ln x \leq x - 1 \quad \text{for all } x > 0. \tag{1.2}$$

Also, we would like to remark that the inequality (1.1) was used to obtain many important results from the foundations of Information Theory such as: the range of the entropy mapping, the Noiseless Coding Theorem, etc. For some recent results which provide similar inequalities see the papers [2-6].

The main aim of this paper is to point out a counterpart inequality for (1.1) and to use it in connection with the *Noiseless Coding Theorem.*

2 The Results

We shall start with the following inequality.

Date. November, 1998
1991 Mathematics Subject Classification. Primary 26D15; Secondary 94Xxx
Key words and phrases. Analytic Inequalities, Noiseless Coding Theorem

Lemma 2.1. *Let p_i, q_i be strictly positive real numbers for $i = 1, ..., n$. Then we have the double inequality:*

$$\frac{1}{\ln r}\sum_{i=1}^{n}(p_i - q_i) \leq \sum_{i=1}^{n}\left(\log_r \frac{1}{q_i} - \log_r \frac{1}{p_i}\right)p_i \leq \frac{1}{\ln r}\sum_{i=1}^{n}\left(\frac{p_i}{q_i} - 1\right)p_i \tag{2.1}$$

where $r > 1, r \in \mathbf{R}$. The equality holds in both inequalities iff $p_i = q_i$ for all i.

Proof. The mapping $f(x) = \log_r x$ is a concave mapping on $(0, \infty)$ and thus satisfies the double inequality

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

for all $x, y > 0$, and as

$$f'(x) = \frac{1}{\ln r}\cdot\frac{1}{x}$$

we get

$$\frac{1}{\ln r}\cdot\frac{x-y}{y} \geq \log_r x - \log_r y \geq \frac{1}{\ln r}\cdot\frac{x-y}{x} \quad \text{for all } x, y > 0. \tag{2.2}$$

Let us choose $x = \frac{1}{q_i}$, $y = \frac{1}{p_i}$ in (2.2) to get

$$\frac{1}{\ln r}\cdot\frac{(p_i - q_i)}{q_i} \geq \log_r \frac{1}{q_i} - \log_r \frac{1}{p_i} \geq \frac{1}{\ln r}\cdot\frac{(p_i - q_i)}{p_i} \tag{2.3}$$

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

Now, if we multiply this inequality by $p_i > 0$ $(i = 1, ..., n)$ we get:

$$\frac{1}{\ln r}\left[p_i\left(\frac{p_i}{q_i} - 1\right)\right] \geq p_i \log_r \frac{1}{q_i} - p_i \log_r \frac{1}{p_i} \geq \frac{1}{\ln r}\cdot(p_i - q_i) \tag{2.4}$$

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

Now, summing over i from 1 to n, we obtain the desired inequality (2.1).

The statement on equality holds by the strict concavity of the mapping $\log_r(.)$. We shall omit the details. ∎

Corollary 2.2. *Let $P = (p_1, ..., p_n)$ be a probability distribution, that is, $p_i \in [0,1]$ and $\sum_{i=1}^{n} p_i = 1$. Let $Q = (q_1, ..., q_n)$ have the property that $q_i \in [0,1]$ and $\sum_{i=1}^{n} q_i \leq 1$(note the inequality here). Then we have:*

$$0 \leq \frac{1}{\ln r}\left(1 - \sum_{i=1}^{n} q_i\right) \leq \sum_{i=1}^{n} p_i \log_r \frac{1}{q_i} - \sum_{i=1}^{n} p_i \log_r \frac{1}{p_i} \leq \frac{1}{\ln r}\left(\sum_{i=1}^{n}\frac{p_i^2}{q_i} - 1\right) \tag{2.5}$$

where $r > 1, r \in \mathbf{R}$. The Equality holds iff $p_i = q_i$ $(i = 1, ...n)$.

The proof is obvious by Lemma 2.1 taking into account that $\sum_{i=1}^{n} p_i = 1$ and $1 \geq \sum_{i=1}^{n} q_i$.

Remark 2.1. *Note that the above corollary is a worth-while improvement of Lemma 1.2.2 from the book [1] which plays there a very important role in obtaining the basically inequalities for entropy, conditional entropy, mutual information, etc.*

Now, consider an encoding scheme $(c_1, ..., c_n)$ for a probability distribution $(p_1, ..., p_n)$. Recall that the *average codeword length* of an encoding scheme $(c_1, ..., c_n)$ for $(p_1, ..., p_n)$ is

$$AveLen(c_1, ..., c_n) = \sum_{i=1}^{n} p_i len\left(c_i\right).$$

We denote the length $len(c_i)$ by l_i .

Recall also that the $r - ary$ entropy of a probability distribution (or of a source) is given by:

$$H_r(p_1, ..., p_n) = \sum_{i=1}^{n} p_i \log_r \frac{1}{p_i}.$$

The following theorem is well known in the literature (see for example [1, Theorem 2.3.1, p. 62]):

Theorem 2.3. *Let $C = (c_1, ..., c_n)$ be an instantaneous (decipherable) encoding scheme for $P = (p_1, ..., p_n)$. Then we have the inequality:*

$$H_r(p_1, ..., p_n) \leq AveLen(c_1, ..., c_n), \tag{2.6}$$

with equality if and only if $l_i = \log_r(\frac{1}{p_i})$ for all $i = 1, ..., n$.

We shall give now the following sharpening of (2.6) which has important consequences in connection with Noiseless Coding Theorem as follows.

Theorem 2.4. *Let C and P be as in the above theorem. Then we have the inequality:*

$$0 \leq \frac{1}{\ln r}\left(1 - \sum_{i=1}^{n} \frac{1}{r^{l_i}}\right) \tag{2.7}$$

$$\leq AveLen\left(c_1, ..., c_n\right) - H_r\left(p_1, ..., p_n\right) \leq \frac{1}{\ln r} \sum_{i=1}^{n} p_i \left(p_i r^{l_i} - 1\right).$$

The Equality holds iff $l_i = \log_r \left(\frac{1}{p_i}\right)$.

Proof. Define $q_i := \frac{1}{r^{l_i}}$ $(i = 1, ...n)$. Then $q_i \in [0, 1]$ and $\sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \frac{1}{r^{l_i}} \leq 1$ by Kraft's theorem (see for example [1, Theorem 2.1.2, p. 44]) and by a simple computation (as in [1, p. 62]) we have :

$$\sum_{i=1}^{n} p_i \log_r \frac{1}{q_i} = \sum_{i=1}^{n} p_i \log_r \left(r^{l_i}\right) = \sum_{i=1}^{n} p_i l_i = AveLen\left(c_1, ...c_n\right).$$

Also

$$\frac{1}{\ln r}\left(\sum_{i=1}^{n}\frac{p_i^2}{q_i}-1\right)=\frac{1}{\ln r}\sum_{i=1}^{n}p\left(r^{l_i}-1\right).$$

Thus inequality (2.5) yields (2.7). ∎

The following theorem also holds.

Theorem 2.5. *Let $P=(p_1,...,p_n)$ be a given probability distribution and $r\in\mathbf{N}$, $r\geq 2$. If $\varepsilon>0$ is given and there exists natural numbers $l_1,...,l_n$ such that*

$$\log_r\left(\frac{1}{p_i}\right)\leq l_i\leq\log_r\left(\frac{1+\varepsilon\ln r}{p_i}\right)\quad\text{for all } i\in\{1,...,n\},\tag{2.8}$$

then there exists an instantaneous r-ary code $C=(c_1,...,c_n)$ with codeword length $len\,(c_i)=l_i$ such that:

$$H_r\left(p_1,...,p_n\right)\leq\ AveLen\left(c_1,...,c_n\right)\leq H_r\left(p_1,...,p_n\right)+\varepsilon.\tag{2.9}$$

Proof. First of all, let us observe that (2.8) is equivalent to

$$\frac{1}{p_i}\leq r^{l_i}\leq\frac{1+\varepsilon\ln r}{p_i},\quad\text{for all } i\in\{1,...,n\}.\tag{2.10}$$

Now, as $\frac{1}{r^{l_i}}\leq p_i$, we deduce that

$$\sum_{i=1}^{n}\frac{1}{r^{l_i}}\leq\sum_{i=1}^{n}p_i=1$$

and by Kraft's theorem, there exists an instantaneous $r-ary$ *code* $C=(c_1,...,c_n)$ so that $len\,(c_i)=l_i$. Obviously, by the Theorem 2.3, the first inequality in (2.9) holds.

We prove the second inequality.

By Theorem 2.4 we have the estimate

$$\begin{aligned}&AveLen\left(c_1,...,c_n\right)-H_r\left(p_1,...,p_n\right)\\&\leq\frac{1}{\ln r}\sum_{i=1}^{n}p\left(p_ir^{l_i}-1\right)\\&\leq\frac{1}{\ln r}\sum_{i=1}^{n}p_i\left|p_ir^{l_i}-1\right|\leq\max_{i=1,\ldots,n}\left\{\left|p_ir^{l_i}-1\right|\right\}\frac{1}{\ln r}\sum_{i=1}^{n}p_i\\&=\frac{1}{\ln r}\max_{i=1,\ldots,n}\left\{\left|p_ir^{l_i}-1\right|\right\}.\end{aligned}\tag{2.11}$$

Now, we observe that (2.10) implies

$$\frac{1-\varepsilon\ln r}{p_i}\leq\frac{1}{p_i}\leq r^{l_i}\leq\frac{1+\varepsilon\ln r}{p_i},\qquad i\in\{1,...,n\},$$

i.e. ,

$$1-\varepsilon\ln r\leq p_ir^{l_i}\leq 1+\varepsilon\ln r,\qquad i\in\{1,...,n\},$$

which is equivalent to

$$\left|p_i r^{l_i} - 1\right| \le \varepsilon \ln r \quad \text{for all } i \in \{1, ..., n\}$$

and then, by (2.11), we deduce the second part of (2.9). ∎

Remark 2.2. *Since for* $\varepsilon \in (0,1)$, *we have for all* $r > 0$,

$$\log_r\left(\frac{1+\varepsilon \ln r}{p_i}\right) - \log_r\left(\frac{1}{p_i}\right) = \log_r(1+\varepsilon \ln r) < \log_r r = 1,$$

(because $1+\varepsilon \ln r < r$ *for all* r *for a given* $\varepsilon \in (0,1)$*) we are not sure always we can find a natural number* l_i *so that inequality* (2.8) *holds.*

Before giving some sufficient conditions for the probability $P = (p_1, ..., p_n)$ so that we can find natural numbers l_i satisfying the inequalities (2.8), let us recall the Noiseless Coding Theorem.

We shall use the notation

$$MinAveLen_r(p_1, ..., p_n)$$

to denote the minimum average codeword length among all $r-ary$ instantaneous encoding scheme for the probability distribution $P = (p_1, ..., p_n)$.

The following Noiseless Coding Theorem is well known in the literature (see for example [1, Theorem 2.3.2, p. 64]) :

Theorem 2.6. *For any probability distribution* $P = (p_1, ..., p_n)$ *we have*

$$H_r(p_1, ..., p_n) \dot{\le} MinAveLen_r(p_1, ..., p_n) < H_r(p_1, ..., p_n) + 1. \tag{2.12}$$

The following question arises naturally:

Question: *Is it possible to replace the constant* 1 *on* (2.12) *by a smaller constant* $\varepsilon \in (0,1)$ *under some conditions on the probability distribution* $P = (p_1, ..., p_n)$ *?*

We are able to give the following (partial) answer to this question.

Theorem 2.7. *Let* r *be a given natural number and* $\varepsilon \in (0,1)$. *If a probability distribution* $P = (p_1, ..., p_n)$ *satisfies the condition that every closed interval*

$$I_i = \left[\log_r\left(\frac{1}{p_i}\right), \log_r\left(\frac{1+\varepsilon \ln r}{p_i}\right)\right], \quad i \in \{1, ..., n\}$$

contains at least one natural number l_i, *then for that probability distribution* P *we have*

$$H_r(p_1, ..., p_n) \le MinAveLen_r(p_1, ..., p_n) \le H_r(p_1, ..., p_n) + \varepsilon. \tag{2.13}$$

Proof. Under the hypotheses

$$\sum_{i=1}^{n} \frac{1}{r^{l_i}} \le \sum_{i=1}^{n} p_i = 1$$

and by Kraft's theorem, there exists an instantaneous code $C = (c_1, ..., c_n)$ so that $len(c_i) = l_i$. For that code we have the condition (2.8) and then, by Theorem 2.5, we have the inequality (2.9) . Taking the infimum in that inequality over all $r - ary$ instantaneous codes, we get (2.13). ∎

The following theorem could be useful for applications.

Theorem 2.8. *Let* a_i $(i = 1, ..., n)$ *be* n *natural numbers. If* p_i $(i = 1, ..., n)$ *are such that*

$$\frac{1}{r^{a_i}} \leq p_i \leq \frac{1 + \varepsilon \ln r}{r^{a_i}} \quad \text{for } i = 1, ..., n; \tag{2.14}$$

and $\sum_{i=1}^{n} p_i = 1$*, then there exists an instantaneous code* $C = (c_1, ..., c_n)$ *with* $len(c_i) = a_i$ *so that* (2.9) *holds for the probability distribution* $P = (p_1, ..., p_n)$*. Furthermore, for that distribution, we have the inequality* (2.13)*.*

Proof. The condition (2.14) is equivalent to

$$\frac{1}{p_i} \leq r^{a_i} \quad \text{and} \quad \frac{1 + \varepsilon \ln r}{p_i} \geq r^{a_i}, \quad i = 1, ..., n;$$

which implies

$$\log_r \left(\frac{1}{p_i} \right) \leq a_i \leq \log_r \left(\frac{1 + \varepsilon \ln r}{p_i} \right), \quad i = 1, ..., n;$$

and then $a_i \in I_i$, $i = 1, ..., n$.

Applying the above results, we get the desired conclusion. ∎

References

[1] S. ROMAN, *Coding and Information Theory*, Springer-Verlag, New York, Berlin, Heidelberg, 1992.

[2] S.S. DRAGOMIR and C.J. GOH, A counterpart of Jensen's discrete inequality for differentiable convex mapping and applications in information theory, *Math. Comput. Modeling,* **24**(2) (1996), 1 - 11 .

[3] S.S. DRAGOMIR and C.J. GOH, Further counterparts of some inequalities in information theory, *submitted.*

[4] S.S. DRAGOMIR and C.J. GOH, A counterpart of Jensen's continuous inequality and applications in information theory, *submitted.*

[5] S.S. DRAGOMIR and C.J. GOH, Some monotonicity and superadditivity properties of the entropy function, *submitted.*

[6] S.S. DRAGOMIR and C.J. GOH, Some bound on entropy measures in information theory, *Appl. Math. Lett.* **10**(3)(1997), 23-28.

School of Communications and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City MC, Victoria 8001, Australia.

E-mail address: {nico, sever}@matilda.vut.edu.au