Some Remarks on the Noiseless Coding Theorem

SOME REMARKS ON THE NOISELESS CODING THEOREM

N.S. BARNETT AND S.S. DRAGOMIR

Abstract. An improvement of the Noiseless Coding Theorem for certain probability distributions is given.

1. Introduction

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

Lemma 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 $\sum_{i=1}^{n} q_i \leq 1$, then*

$$\sum_{i=1}^{n} p_i \log_b \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log_b \frac{1}{q_i} \quad (b > 1) \tag{1.1}$$

where $0 \log_b \frac{1}{0} = 0$ and $p \log_b \frac{1}{0} = +\infty$ for $p > 0$. Furthermore, the equality holds if and only if $q_i = p_i$ for all i.

Note that the proof of this result in [1] uses the elementary inequality:

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

We give here an alternative proof based on the concavity of the mapping $\log_r(\cdot)$.

As the mapping $f(x) = \log_r(x)$ $(r > 1)$ is a strictly concave mapping on $(0, \infty)$, we have

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

for all $x, y > 0$, i.e., as $f'(x) = \frac{1}{\ln r} \cdot \frac{1}{x}$ for $x > 0$,

$$\log_r x - \log_r y \geq \frac{1}{\ln r}\left(\frac{x - y}{x}\right) \tag{1.2}$$

for all $x, y > 0$.

Choosing $x = \frac{1}{q_i}, y = \frac{1}{p_i}$, in (1.2) gives

$$\log_r \frac{1}{q_i} - \log_r \frac{1}{p_i} \geq \frac{1}{\ln r}\left(\frac{p_i - q_i}{p_i}\right) \tag{1.3}$$

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

Multiplying this inequality by $p_i > 0$ $(i = 1, ..., n)$ we get

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

Date: February 26, 2004.

1991 *Mathematics Subject Classification.* Primary 94 A xx; Secondary 26 D xx.

Key words and phrases. Entropy, Instantaneous Codes, The Noiseless Coding Theorem.

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

Summing over i from 1 to n, gives

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

$$= \frac{1}{\ln r} \left(1 - \sum_{i=1}^{n} q_i \right) \geq 0$$

and the inequality (1.1) is obtained.

The case of equality follows by the strict concavity of the mapping log_r.

In this paper, by use of (1.1), we point out an improvement to the Noiseless Coding Theorem.

2. The Results

Consider an encoding scheme $(c_1, ..., c_n)$ for a probability distribution $(p_1, ..., p_n)$. The *average codeword length of an encoding scheme* $(c_1, ..., c_n)$ for $(p_1, ..., p_n)$ is

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

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

The $r-$ary entropy of a probability distribution is given by

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

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

Theorem 2. *Let $C = (c_1, ..., c_n)$ be an instantaneous (or uniquely decipherable) encoding scheme for $P = (p_1, ..., p_n)$, then,*

$$H_r(p_1, ..., p_n) \leq AveLen(c_1, ..., c_n)$$

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

The following result, providing a counterpart inequality, holds.

Theorem 3. *Let $P = (p_1, ..., p_n)$ be a given probability distribution and $r \in \mathbf{N}, r \geq 2$. If $\varepsilon > 0$ is fixed and there exists natural numbers $l_1, ..., l_n$ such that:*

$$\log_r \left(\frac{1}{p_i} \right) \leq l_i \leq \log_r \left(\frac{r^{\varepsilon}}{p_i} \right) \tag{2.1}$$

for all $i \in \{1, ..., n\}$, then there exists an instantaneous $r-$ary code $C = (c_1, ..., c_n)$ with codeword length $len(c_i) = l_i$ such that

$$H_r(p_1, ..., p_n) \leq AveLen(c_1, ..., c_n) \leq H_r(p_1, ..., p_n) + \varepsilon. \tag{2.2}$$

Proof. Note that (2.1) is equivalent to

$$\frac{1}{p_i} \leq r^{l_i} \leq \frac{r^{\varepsilon}}{p_i} \quad \text{for all } i \in \{1, ..., n\}. \tag{2.3}$$

Now, since $\frac{1}{r^{l_i}} \leq p_i$ $(i = 1, ..., n)$, it follows that

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

and by Kraft's theorem (see for example [1, Theorem 2.1.2, p. 44]), there exists an instantaneous $r-$ary code $C = (c_1, ..., c_n)$ such that $len(c_i) = l_i$.

Obviously, by Theorem 2, the first inequality in (2.2) holds.

We have:

$$\begin{aligned} &AveLen(c_1, ..., c_n) \\ &= \sum_{i=1}^{n} p_i l_i = \sum_{i=1}^{n} p_i \log_r r^{l_i} = \sum_{i=1}^{n} p_i \log_r \frac{1}{q_i} \end{aligned}$$

choosing $q_i = \frac{1}{r^{l_i}} \in [0, 1]$. Also, by Kraft's theorem, $\sum_{i=1}^{n} q_i \leq 1$.

By Lemma 1, we have,

$$\begin{aligned} 0 \leq \sum_{i=1}^{n} p_i \log_r \frac{1}{q_i} - \sum_{i=1}^{n} p_i \log_r \frac{1}{p_i} &= AveLen(c_1, ..., c_n) - H_r(p_1, ..., p_n) \\ = \sum_{i=1}^{n} p_i \left(\log_r r^{l_i} - \log_r \frac{1}{p_i} \right) &= \left| \sum_{i=1}^{n} p_i \left(\log_r r^{l_i} - \log_r \frac{1}{p_i} \right) \right| \\ \leq \sum_{i=1}^{n} p_i \left| l_i - \log_r \left(\frac{1}{p_i} \right) \right| &\leq \varepsilon \sum_{i=1}^{n} p_i = \varepsilon \end{aligned}$$

since, by (2.1), $0 \leq l_i - \log_r \frac{1}{p_i} \leq \log_r r^{\varepsilon} = \varepsilon$. ∎

We shall use the notation:

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

to denote the minimum average codeword length among all $r-$ary instantaneous encoding schemes 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 4. *For any probability distribution $P = (p_1, ..., p_n)$ we have:*

$$H_r(p_1, ..., p_n) \leq MinAveLen_r(p_1, ..., p_n) < H_r(p_1, ..., p_n) + 1. \tag{2.4}$$

The following question is then a natural one to pose.

Question: *Is it possible to replace the constant* 1 *in the above inequality by a smaller one $\varepsilon \in (0, 1)$ and, if so, under what conditions for the probability distribution $P = (p_1, ..., p_n)$?*

The following is a partial answer to this question:

Theorem 5. *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 of real numbers*

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

contains one natural number, then, for that probability distribution P, we have:

$$H_r(p_1, ..., p_n) \leq MinAveLen_r(p_1, ..., p_n) \leq H_r(p_1, ..., p_n) + \varepsilon. \tag{2.5}$$

Proof. Suppose that $l_i \in I_i$ $(i = 1, ..., n)$ are these natural numbers, then, as above,

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

and by Kraft's theorem there exists an instantaneous code $C = (c_1, ..., c_n)$ such that $len(c_i) = l_i$. For this code we have (2.1) and, by Theorem 3, the inequality (2.2) for C. Taking the infimum in this inequality over all $r-$ary instantaneous codes, gives (2.5). ∎

Remark 1. *The lengths of the intervals I_i are,*

$$len(I_i) = \log_r\left(\frac{r^{\varepsilon}}{p_i}\right) - \log_r \frac{1}{p_i} = \varepsilon \in (0, 1), \qquad i = 0, ..., n$$

but we cannot be sure that I_i always contains a natural number. Also, I_i could contain at most one natural number.

The following result can be useful in practice.

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

$$\frac{1}{r^{a_i}} \leq p_i \leq \frac{r^{\varepsilon}}{r^{a_i}} \qquad \text{for } i = 1, ..., n \tag{2.6}$$

and $\sum_{i=1}^{n} p_i = 1$, then there exists an instantaneous code $C = (c_1, ..., c_n)$ with $len(c_i) = a_i$ $(i = 1, ..., n)$ such that (2.2) holds for the probability distribution $P = (p_1, ..., p_n)$.

For other recent results in the applications of Theory of Inequalities in Information Theory and Coding, see the following references.

References

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

[2] N.M. Dragomir and S.S Dragomir, An inequality for logarithms and its application in coding theory, *Indian J. Math.*, **43**(1) (2001), 13-20.

[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.

[7] S.S. Dragomir, N.M. Dragomir and K. Pranesh, Some Estimations of Kraft number and related results, *Soochow Journal of Mathematics*, **24**(4)(1998), 291-296.

[8] S.S. Dragomir, M. Sholz and J. Sunde, Some upper bounds for relative entropy and applications, *Comp. & Math. with Applications*, **39**(2000), 91-100.

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

School of Computer Science & Mathematics, Victoria University of Technology, PO Box 14428, Melbourne City MC, Victoria 8001, Australia.

E-mail address: {neil,sever}@matilda.vu.edu.au