Kraft's Number and Ideal Word Packing

N. M. Dragomir*, S. S. Dragomir*, C. E. M. Pearce† and J. Šunde‡

*School of Communications and Informatics,
Victoria University of Technology, PO Box 14428, MCMC, Melbourne, Victoria
†Applied Mathematics Department
The University of Adelaide, Adelaide SA 5005
‡Communication Division, DSTO, PO Box 1500 Salisbury, SA 5108[1]

Abstract. In the noiseless context, it has long been known that the average encoded word length of an instantaneous or uniquely decipherable code can be made to lie between the source entropy and that value plus unity. We address the question of finding sufficient conditions on the code–word probabilities for it to be possible to make the average code–word length approximate the entropy by a smaller prescribed amount.

INTRODUCTION

The nomenclature *noisy coding* and *noiseless coding* tends to obscure relations between the two, both of which in fact exhibit order–disorder phenomena. The central role of entropy in both is testimony to this. An analogy may be drawn with molecular behaviour in solid and fluid phases. Disorder in the former manifests itself in such aspects as crystal–packing problems, which correspond to *word packing* in noiseless coding.

Suppose there are n code words, the i–th having length ℓ_i letters drawn from an alphabet of r letters, and suppose that the i–th word occurs with probability p_i. Then the average code word length is $\overline{\ell} = \sum_{i=1}^{n} p_i \ell_i$ and the r–ary entropy or uncertainty of the message source is

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

where, as subsequently, $\log_r$ refers to logarithms taken to base r. The following theorem is well–known in the literature (see, for example, [1, p. 62]).

Theorem A. *In an instantaneous or uniquely decipherable code we have*

[1] Acknowledgement. This work was supported financially by DSTO.

CP511, *Unsolved Problems of Noise and Fluctuations*, edited by D. Abbott and L. B. Kish

$$H_r \leq \bar{\ell}, \tag{1}$$

with equality if and only if $\ell_i = \log_r 1/p_i$ *for all* $i = 1, ..., n$.

Here equality reflects ideal word packing in which the theoretical minimum value, the entropy of the word source, is achieved with optimal word packing. For efficient transmission we desire that the difference between the two sides of (1) be as small as possible.

We adopt the notation $A_r(p_1, \ldots, p_n)$ for the minimum average codeword length over all r–ary instantaneous encoding schemes for the probability distribution $(p_i)_{i=1}^n$. The *noiseless coding theorem* (see for example [1, Theorem 2.3.2, p. 64]) states the following.

Theorem B. *For any probability distribution* $(p_i)_{i=1}^n$ *we have*

$$H_r(p_1, ..., p_n) \leq A_r(p_1, ..., p_n) < H_r(p_1, ..., p_n) + 1. \tag{2}$$

With the use of coding in blocks of k words, this relation can be improved to

$$H_r(p_1, \ldots, p_n) \leq A_r^{(k)}(p_1, \ldots, p_n) < H_r(p_1, \ldots, p_n) + 1/k, \tag{3}$$

where $A_r^{(k)}$ refers to the average length per word when the coding is effected in blocks of k.

These are generic results that take no account of any special structure that may exist in a coding situation. By analogy with crystal packing, we would expect that special structure could be used to improve the result. This leads to the following.

Question 1. Under what conditions can the additive constant unity in (2) be replaced by a given $\epsilon \in (0, 1)$ without the use of block coding?

A natural parameter in the examination of such questions is Kraft's number, which is defined by

$$K_r := \sum_{i=1}^{n} r^{-\ell_i}.$$

Thus by a result due to Kraft and McMillan (see [2, Chapter 2]) the condition $K_r \leq 1$ (known as Kraft's inequality) is necessary and sufficient for the existence for the existence of an instantaneous code with code words of lengths ℓ_i $(1 \leq i \leq n)$.

Some preliminary ideas have been developed in [3], where the authors have derived the following improvement of (1).

Theorem C. *For an instantaneous code,*

$$0 \leq \frac{1}{\ln r}(1 - K_r) \leq \bar{\ell} - H_r \leq \frac{1}{\ln r} \sum_{i=1}^{n} p_i \left(p_i r^{\ell_i} - 1 \right). \tag{4}$$

Equality holds if and only if $\ell_i = \log_r(1/p_i)$.

In this paper we extend this work. In particular, we address the question of finding sufficient conditions on the distribution $(p_i)_{i=1}^n$ for the difference between the two sides of (1) to be no more than ϵ for a given $\epsilon \in (0,1)$.

In the following section we present a basic mathematical tool useful for deriving such results as Theorem C. This we then use in the subsequent section for addressing Question 1.

BASIC INEQUALITY

Proposition 1. *Suppose r is an integer exceeding unity and p_i, q_i are strictly positive real numbers ($i = 1, ..., n$). Then we have the double inequality*

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

Both inequalities reduce to equality if and only if $p_i = q_i$ for each i.

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

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

for all positive x, y. Since $f'(x) = 1/(x \ln r)$, we have

$$\frac{1}{\ln r}\frac{x - y}{y} \geq \log_r x - \log_r y \geq \frac{1}{\ln r}\frac{x - y}{x}$$

for all $x, y > 0$.

The choices $x = 1/q_i$ and $y = 1/p_i$ provide

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

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

Multiplication by p_i and summing over i yields the desired inequalities. The statement on equality follows from the strict concavity of the mapping $\log_r(\cdot)$. □

In particular, if $(p_i)_{i=1}^n$, $(q_i)_{i=1}^n$ are probability distributions, then we have

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

with equality if and only if $p_i = q_i$ $(1 \leq i \leq n)$.

This is a refinement of the fundamental lemma of information theory.

BOUNDS

McMillan has shown that for there to exist an instantantaneous or uniquely decipherable r–ary code with code words of lengths ℓ_i, Kraft's inequality $K_r \leq 1$ must be satisfied (see [1, pp. 47–49]). This is, of course, implicit in the first inequality in (4). The following result therefore gives nontrivial upper and lower bounds for the difference between the two sides of (1).

Theorem 1. *For an instantaneous code,*

$$0 \leq \log_r K_r + \bar{\ell} - H_r \leq \frac{1}{\ln r}\left[K_r \sum_{i=1}^{n} p_i^2 r^{\ell_i} - 1\right], \tag{5}$$

with equality if and only if

$$p_i = r^{-\ell_i}/K_r \tag{6}$$

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

Proof. The choice $q_i := r^{-\ell_i}/K_r$ in Proposition 1 yields

$$\begin{aligned} 0 &= \frac{1}{\ln r}\left(1 - K_r^{-1} \sum_{i=1}^{n} r^{-\ell_i}\right) \\ &\leq \sum_{i=1}^{n} p_i \log_r \left(K_r r^{\ell_i}\right) - \sum_{i=1}^{n} p_i \log_r 1/p_i \\ &\leq \frac{1}{\ln r}\left(K_r \sum_{i=1}^{n} p_i^2 r^{\ell_i} - 1\right), \end{aligned}$$

which is equivalent to (5). The statement about equality follows directly from that in Proposition 1. □

The following result holds also.

Theorem 2. *Suppose* $(p_i)_{i=1}^n$ *is a probability distribution and* $(c_i)_{i=1}^n$ *a set of code words,* c_i *occurring with probability* p_i. *Let* $r \geq 2$ *be a positive integer. If* $\epsilon > 0$ *is fixed and there exist positive integers* $\ell_1, \ldots, \ell_n$ *such that*

$$\log_r(1/p_i) \leq \ell_i \leq \log_r(r^\epsilon/p_i) \tag{7}$$

for all $i \in \{1, ..., n\}$, *then there exists an instantaneous* r*–ary code in which* c_i *has length* ℓ_i *and*

$$H_r \leq \bar{\ell} \leq H_r + \epsilon. \tag{8}$$

Proof. We rewrite (7) as

$$1/p_i \leq r^{\ell_i} \leq r^{\epsilon}/p_i$$

$(1 \leq i \leq n)$. Since $r^{-\ell_i} \leq p_i$, we deduce that

$$K_r = \sum_{i=1}^{n} r^{-\ell_i} \leq \sum_{i=1}^{n} p_i = 1.$$

By Kraft's Theorem [1, Theorem 2.1.2, p. 44] there exists an instantaneous r–ary code in which c_i has length ℓ_i.

The first inequality in (8) holds by Theorem A. By choosing $q_i = r^{-\ell_i} \in (0,1)$ we have

$$\overline{\ell} = \sum_{i=1}^{n} p_i \log_r r^{\ell_i} = \sum_{i=1}^{n} p_i \log_r (1/q_i)$$

and $\sum_{i=1}^{n} q_i \leq 1$. Further by Proposition 1

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} p_i \log_r \frac{1}{q_i} - \sum_{i=1}^{n} p_i \log_r \frac{1}{p_i} \\ &= \sum_{i=1}^{n} p_i \left(\ell_i - \log_r \frac{1}{p_i} \right) \\ &\leq \epsilon \sum_{i=1}^{n} p_i \\ &= \epsilon, \end{aligned}$$

since, by (7), we have that $0 \leq \ell_i - \log_r(1/p_i) \leq \log_r r^{\epsilon} = \epsilon$. □

We may now give a partial answer to Question 1.

Theorem 3. *Let $r \geq 2$ be a positive integer and $\epsilon \in (0,1)$. If the probability distribution $(p_i)_{i=1}^{n}$ satisfies the condition that every closed interval*

$$I_i := [\log_r(1/p_i), \log_r(r^{\epsilon}/p_i)], \quad i \in \{1, ..., n\}$$

contains an integer, then

$$H_r(p_1, ..., p_n) \leq A_r(p_1, ..., p_n) \leq H_r(p_1, ..., p_n) + \epsilon. \tag{9}$$

Proof. Suppose that $l_i \in I_i$ $(i = 1, ..., n)$ are these integers. Then

$$K_r = \sum_{i=1}^{n} r^{-\ell_i} \leq \sum_{i=1}^{n} p_i = 1$$

and by Kraft's theorem there exists an instantaneous code in which c_i has length ℓ_i. For that code we have condition (7) and by Theorem 2, (8) holds. Taking the infinum over all r–ary instantaneous codes yields (9). □

The following reformulation of Theorem 2 can be useful in practice.

Practical Criterion: Let $(a_i)_{i=1}^n$ be positive integers and $(p_i)_{i=1}^n$ such that

$$r^{-a_i} \leq p_i \leq r^{\epsilon - a_i}$$

for all $i = 1, ..., n$ and $\sum_{i=1}^n p_i = 1$. Then there exists an instantaneous code in which c_i has length a_i such that (8) and (9) hold.

CONCLUSION & OPEN QUESTIONS

We have found conditions under which the constant unity in (2) can be replaced by a given $\epsilon \in (0, 1)$ and recast these in terms of a practical criterion. In practice block coding is usually employed. So a natural follow–up to our work is the following.

Question 2. Do our results have manageable extensions in block coding in which the final term in (3) can be replaced by ϵ/k for given $\epsilon \in (0, 1)$?

Kraft's inequality, or rather its extension to (infinite) recursively enumerable sets, the Kraft–Chaitin inequality ([4], [5], see also [6]) has important ongoing ramifications in connection with the theory of program size, which Chaitin has shown to have a structure identical to information theory. Accordingly we may also foreshadow a rather more general question.

Question 3. What implications do our discussion in the previous section have for the Kraft–Chaitin inequality and the theory of program size?

REFERENCES

1. Roman, S., *Coding and Information Theory*, Heidelberg: Springer Verlag, 1992.
2. Ash, R., *Information Theory*, New York: Interscience, 1967.
3. Dragomir, S. S. and Dragomir, N. M., An inequality for logarithms and its applications in coding theory, *Indian J. Math.* (in press).
4. Chaitin, G. J., A theory of program size formally identical to information theory, *J. Assoc. Comput. Mach.* **22**, 329–340 (1975).
5. Calude, C. and Grozea, C., Kraft–Chaitin inequality revisited, *J. of Universal Computer Science*, **5**, 306–310 (1996).
6. Chaitin, G. J., Information–theoretic characterizations of recursive infinite strings, *Theoret. Comput. Sci.* **2**, 45–48 (1976).