A BLOCK CODING THEOREM AND RÉNYI'S ENTROPY

L. LORNE CAMPBELL

ABSTRACT. It is shown that some recent results of Csiszár on block coding for sources can be derived from an earlier result on variable-length coding.

1. INTRODUCTION

Some time ago, the author [3] examined the problem of finding an optimal variable-length code for a discrete memoryless source when the cost of a code-word is an exponential function of its length. An exponential cost function can be more appropriate than a linear function if the processing cost for decoding is significant or if buffer overflow caused by long codewords is important. For further discussion of exponential cost functions, see Humblet [7], Blumer and McEliece [2], the author [4], and references in these papers. A unified and extended treatment of the material in [3] and [4] is given by Aczél and Daróczy [1, pp.156–172]. For exponential costs, the Rényi entropy plays a role in the coding theorem that is similar to the role of the Shannon entropy for a linear cost function.

In an interesting recent paper, Csiszár [6] examines a different kind of coding problem for the discrete memoryless source. He is concerned with block codes whose probability of error has a given exponential rate of convergence to zero with block length. He finds that the code rates for which the error probability decreases exponentially to zero are bounded below by Rényi entropy of an order which is fixed by the desired exponential rate of convergence to zero. (He also gives similar results for noisy channel coding and for hypothesis testing, but these results are not directly relevant to this note.)

Date: January 21, 2000.

1991 *Mathematics Subject Classification.* 94A29.

Key words and phrases. Source coding, Rényi entropy.

This research was supported by the Natural Sciences and Engineering Research Council of Canada, Grant Number OGP0002151.

The object of this note is to use the variable-length result to give a different derivation of some of Csiszár's [6] results. This may help to clarify the relationship between the two kinds of coding theorem.

2. Definitions and Notation

Let the source produce symbols from the alphabet of size N, $\mathcal{X} = \{x_1, \dots, x_N\}$ with probabilities $P(x_i) = p_i$, where $\sum p_i = 1$, and let symbols be produced independently, so that

$$P(x_{i_1} x_{i_2} \cdots x_{i_n}) = p_{i_1} p_{i_2} \cdots p_{i_n}.$$

Let there be a code alphabet $\mathcal{A}$, of size D, which will be used to encode the source. Define the Rényi entropy of order α ($\alpha \neq 1$) of this source to be

$$H_\alpha(\mathcal{X}) = \frac{1}{1-\alpha} \log_D \left(\sum_{i=1}^{N} (p_i)^\alpha \right). \tag{1}$$

Because successive symbols are independent, it is easy to show that

$$H_\alpha(\mathcal{X}^n) = n H_\alpha(\mathcal{X}). \tag{2}$$

Recall that a *variable-length* uniquely decodable code of $\mathcal{X}$ into $\mathcal{A}$ is a map which represents each $x_i \in \mathcal{X}$ by a finite sequence (the code word $w(x_i)$) of length l_i of elements of $\mathcal{A}$ and which has the property that every concatenation $w(x_{i_1}) \cdots w(x_{i_n})$ can be decoded in only one way to a source sequence $x_{i_1} x_{i_2} \cdots x_{i_n}$. It is a well-known result [5] that the lengths l_i in a uniquely decodable code satisfy the Kraft inequality

$$\sum_{i=1}^{N} D^{-l_i} \leq 1 \tag{3}$$

and that, conversely, if $\{l_i, i = 1, \dots, n\}$ are integers which satisfy (3) then there is a uniquely decodable code with these values as lengths of code words. This code can be chosen to have the *prefix property*, i.e. the property that no code word is a prefix of another code word.

It is shown in [4] that for any uniquely decodable code and any $\beta > -1$, except $\beta = 0$,

$$L(\beta) \geq H_\alpha(\mathcal{X}), \tag{4}$$

where $\alpha = (1+\beta)^{-1}$ and $L(\beta)$ is defined by

$$L(\beta) = \frac{1}{\beta} \log_D \left(\sum_{i=1}^{N} p_i D^{\beta l_i} \right).$$

By a *block code* of length n and rate R we mean a map f (the encoding function) from $\mathcal{X}^n$ into $\mathcal{A}^m$ where $R = m/n$ together with a decoding function $g : \mathcal{A}^m \to \mathcal{X}^n$. The probability of error of the code, $P_e(n, R)$, is the probability that the source produces a sequence $x = (x_{i_1} x_{i_2} \cdots x_{i_n})$ which is such that $g(f(x)) \neq x$. Note that if $R < \log_D N$ then $P_e(n, R) > 0$ and the code is not uniquely decodable.

3. Coding Theorems

We now prove the following theorem which was proved by Csiszár by another method:

Theorem 1. *Let $\beta > 0$. There exists a block code of length n and rate R with*

$$P_e(n, R) < D^{-n\beta(R - H_\alpha) + \beta}, \tag{5}$$

where $\alpha = (1+\beta)^{-1}$ and $H_\alpha = H_\alpha(\mathcal{X})$.

Proof. Consider a variable-length code with the prefix property from $\mathcal{X}^n$ into code words from $\mathcal{A}$. Convert this into a block code of length n and rate R by truncating all code words of length greater than m by deleting symbols and by adding extra symbols at the end to all code words with lengths less than m (where, as before, $R = m/n$). With an obvious choice of decoding function, this code makes a decoding error only if the original code word (before truncation) has length greater than m.

We estimate $P_e(n, R)$ by a Chernoff bound argument. Denote by $|x|$ the length of the variable-length code word for $x \in \mathcal{X}^n$ and let $\phi(\beta)$ be the generating function

$$\phi(\beta) = \sum_x P(x) D^{\beta |x|}.$$

Then,

$$\begin{aligned} P_e(n, R) &= P\{|x| > m\} \\ &< \sum_{|x| > m} P(x) D^{\beta(|x| - m)} \le \phi(\beta) D^{-n\beta R}. \end{aligned} \tag{6}$$

But, it follows from [3] that it is possible to choose the variable-length code so that

$$\phi(\beta) < D^{n\beta H_\alpha + \beta},$$

where $\alpha = (1+\beta)^{-1}$. When this code is the one which is truncated, it follows that

$$P_e(n, R) < D^{-n\beta(R - H_\alpha) + \beta}. \qquad \square$$

In the terminology of Csiszár [6], (5) shows that H_α is a β-achievable rate if $\alpha = (1+\beta)^{-1}$. We also remark that (4), applied to codes on $\mathcal{X}^n$ instead of on $\mathcal{X}$, together with (2) shows that

$$\phi(\beta) \geq D^{n\beta H_\alpha}$$

for every possible code. This means that $D^{-n\beta(R-H_\alpha)}$ is the tightest possible bound of the Chernoff type (6). Csiszár [6] shows that this is the best possible bound of its type which can be obtained by any method of bounding and thus that H_α is what he terms the β-cutoff rate.

There is a similar theorem for $\beta < 0$:

Theorem 2. *Let $-1 < \beta < 0$. For any block code of length n and rate R,*

$$P_e(n, R) \geq 1 - D^{n\beta(H_\alpha - R)}, \tag{7}$$

where $\alpha = (1+\beta)^{-1}$ and $H_\alpha = H_\alpha(\mathcal{X})$.

Proof. For any block code of length n and rate R, let $\mathcal{C}$ be the set of source sequences $x = (x_{i_1} x_{i_2} \cdots x_{i_n})$ which are decoded correctly. The block code can be converted into a uniquely decodable code by adjoining a code word of suitable length greater than m for each source sequence in the complement of $\mathcal{C}$ in $\mathcal{X}^n$. Then, since β is negative,

$$\begin{aligned} 1 - P_e(n, R) &= \sum_{|x| \leq m} P(x) \\ &\leq \sum_{|x| \leq m} P(x) D^{\beta(|x|-m)} \leq \phi(\beta) D^{-n\beta R}. \end{aligned} \tag{8}$$

From (4) and (2) and the negativity of β, it follows that

$$\phi(\beta) \leq D^{n\beta H_\alpha}.$$

The theorem follows from this and (8). □

In the terminology of Csiszár [6], (7) shows that H_α is a β-unachievable rate if $\alpha = (1+\beta)^{-1}$. Moreover, H_α can be seen to be the largest β-unachievable rate which can be deduced by an argument of this type. Csiszár [6] has shown that it is the largest which can be found by any method and so in his terminology, H_α is the β-cutoff rate.

References

[1] J. Aczél and Z. Daróczy, *Measures of Information and Their Characterizations*, Academic Press, New York, 1975.
[2] A. C. Blumer and R. J. McEliece, *The Rényi redundancy of generalized Huffman codes*, IEEE Trans. Inform. Theory **34** (1988), 1242–1249.
[3] L. L. Campbell, *A coding theorem and Rényi's entropy*, Information and Control **8** (1965), 423–429.
[4] ———, *Definition of entropy by means of a coding problem*, Z. Wahrscheinlichkeitstheorie verw. Geb. **6** (1966), 113–118.
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
[6] I. Csiszár, *Generalized cutoff rates and Rényi's information measures*, IEEE Trans. Inform. Theory **41** (1995), 26–34.
[7] P. A. Humblet, *Generalization of Huffman coding to minimize the probability of buffer overflow*, IEEE Trans. Inform. Theory **IT-27** (1981), 230–232.

Department of Mathematics and Statistics, Queen's University, Kingston, Ontario Canada K7L 3N6
E-mail address: `campblll@llc.mast.queensu.ca`