J. Indian Inst. Sci., March–April 1997, 77, 209–215.


# Bound on exponential mean codeword length of size $d$ – alphabet 1:1 code*

D. S. HOODA
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar 125 004 India.

**Abstract**

In the present communication we define the codes which assign $d$ – alphabet one-one codeword to each of a random variable and the functions which represent possible transformations from codeword length of a non-one code to codeword length of a uniquely decodable code. By using these functions we obtain bounds on the exponentiated mean codeword length for one-one code of size $d$ – alphabet in terms of Renyi entropy and study the particular cases also.

**Keywords:** One-one code, uniquely decodable code, codeword length and Renyi entropy.

Mathematical Subject Classification (1980) : 94 A 15 and 94 A 24.

## 1. Introduction

Shannon proved that the minimal expected codeword length L of a prefix code for a random variable $X$ satisfies.

$$H(X) \leq L_{UD} < H(X) + 1 \tag{1.1}$$

where $H(X)$ is the Shannon entropy of the random variable $X$ and $L_{UD} = \Sigma p_i l_i$ is the average codeword length for uniquely decodable code. Shannon's restriction that encoding of $X$ will be concatenated and must be uniquely decodable is motivated by the desire to deal with sequential data. In some situations it is advantageous to transmit a single random variable in stead of a sequence of random variables, particularly, when there are $N$ states for one memoryless source, one for each symbol $s_i$ of the source alphabet. Such codes which assign a distinct binary codeword to each outcome of the random variable without regard to the constraint that concatenation of these description be uniquely decodable are called 1:1 codes.

Leung-Yan-Cheong and Cover [4] considered 1:1 codes and defined the average codeword length for the best 1:1 code and obtained its lower bound given as

$$L_{1:1} = \Sigma p_i[\log(i/2 + 1)] \geq H(X) - \log \text{longN} - 3, \tag{1.2}$$

where $N$ is the number of values that random variable $X$ can have and [S] denotes the smallest integer greater than or equal to S. Since the class of 1:1 codes contains the class of uniquely decodable codes, therefore it follows that

*This paper was prepared when the author was with the Universitat Mannheim under Indo-German Cultural Exchange Programme.

$$L_{1:1} \leq L_{UD} \tag{1.3}$$

It may be noted that all logarithms have been taken to the base D unless otherwise stated and we denote the average codeword length for the best 1:1 codes and uniquely decodable codes by $L_{1:1}$ and $L_{UD}$ respectively.

Compbell [1] introduced the exponentiated codeword length.

$$L_{UD}(t) = (1/t)\log\,(\Sigma p_i D^{tl_i}),\ 0 < t < \infty \tag{1.4}$$

where $L_{UD}(t)$ is the average codeword length for uniquely decodable code, D represent the size of code alphabet and $l_i$ is the codeword length associated with $x_i$ of $x$. He proved the following noiseless coding theorem:

$$H_\alpha(X) \leq L_{UD}(t) < H_\alpha(X) + 1 \tag{1.5}$$

$$\text{under the condition } \Sigma D^{-l_i} \leq 1 \tag{1.6}$$

where $H_\alpha(X)$ is Renyi [5] entropy of order $\alpha = 1/1 + t$ and $l_i$ is the codeword length corresponding to source symbols $x_i$. The inequality (1.6) is Kraft's inequality which is necessary and sufficient for the existence of uniquely decodable code.

Kiefer [3] defined a class of decision rules and showed that $H_\alpha(X)$ is the best decision rule for deciding which of two sources can be coded with small expected cost for sequence of length $n$, as $n \to \infty$, where the cost of encoding a sequence is assumed to be a function only of the codeworth length. Jelinek [2] showed that coding with respect to $L_{UD}(t)$ is useful in minimizing the problem of buffer overflow which occurs when the source symbols are being produced at a fixed rate, and the codewords are stored temporarily in a finite buffer.

In the present paper we define the codes which assign D alphabet one to one codeword to each outcome of a random variable and the functions which represent possible transformations from codeword lengths of 1:1 code to UD codes of size D alphabet in section 2.

In section 3 by using these functions we obtain bounds on the exponentiated mean codeword length of the best 1:1 code of size D alphabet in terms of Renyi entropy and study the particular case also.

## 2. Transformation from Codeword lengths of 1:1 to UD Codes of Size D-Alphabet

Let $X = \{x_1, x_2 ..., x_N\}$ be a random variable with finite number of values having discrete probability distribution $P = \{(p_1, p_2,..., p_N)\}$, $p_i > 0$ for all $i$, $\Sigma p_i = i$, $p_i \geq p_j$ for $i < j\}$. Let $1_i$, $i = 1, 2,..., N$, be the codeword length of the sequence encoding $x_i$ in the best 1:1 code of size $D$ alphabet, then the set of possible codewords is

$$\{0, 1,..., (D-1); 00, 01,..., (D-1)\,(D-1); 000, 001,....\}$$

and consequently, we have

$$l_1 = 1,\ l_2 = 1,...,\ l_D = 1,\ l_{D+1} = 2,...,\ l_{D(D+1)} = 2,...\ \text{etc}$$

Thus by inspection we can see that

$$l_i = [\log(D-1)i/D + 1)] \tag{2.1}$$

where $[S]$ denotes the smallest integer greater than or equal to $S$.

Now we define a function $h(l_i)$ such that $\Sigma D^{-h(li)} \leq 1$ holds, only then the set of length $\{h(l_i)\}$ yields acceptable codeword length for a uniquely decodable code. Evidently, if $h$ is an integer valued function such that $\sum_{i=1}^{N} D^{-h(li)} > 1$, then $\{h(l_i)\}$ cannot yield a uniquely decodable code.

Theorem 1. The following functions are possible transformations from codeword lengths of 1:1 codes to those of uniquely decodable codes of size $D$ alphabet:

(i)
$$h(l_i) = l_i + a[\log l_i] + \log (D^a - 1)/ (D^a - D). \text{ where } a > 1, D \geq 2. \tag{2.2}$$

(ii)
$$h(l_i) = l_i + a[\log(l_i + 1)]. \ a > 2 \text{ and } D \geq 2. \tag{2.3}$$

(iii)
$$h(l_i) = l_i + [\log l_i + \log(\log l_i) + \ldots\ldots + \log(\log(\ldots(\log l_i)))] + 4. \tag{2.4}$$

where we only take the first $K$ iterates for which

$$\log(\log(\ldots(\log l_i))) \text{ is positive.}$$

For the proof refer to Leung-Yan-Cheong and Cover[4] Theorem 2.

Lemma 2.1 Let

$$G_D(X) = 1 \times \log_D \times \log_D(\log_D x) \tag{2.8}$$

then

$$I_D = G_D(x) = \int_1^a dx / x \log_D X \log(\log_D X)\ldots.. = \begin{bmatrix} \text{infinite if } D \geq e \\ \text{finite if } D < e \end{bmatrix}$$

For proof refer to [4]

Thus $I_D$ is finite only if $D < e$ which implies $D = 2$.

for $D = 2$ and $M = \text{Log}_2 e$, we have.

$$I_2 \leq \log_2 e/(\log_2 e - 1) < 3.26 \tag{2.9}$$

Hence

$$\Sigma\ 1/l \log l \log(\log 1)\ldots < I_2 + 1 < 5. \tag{2.10}$$

Substituting (2.10) in (2.7), we get

$$S < 5.2^{-c+1} \tag{2.11}$$

If we choose $c = 4$ in (2.11), then $S \leq 1$. Hence theorem 1 is proved.

## 3. Bounds on exponential mean codeword length of 1:1 code

The exponentiated mean codeword length of size $D$ alphabet for the best 1:1 code is defined as

$$L_{1:1}(t) = (1/t)\log\left(\sum p_i D^{t[\log(D-1)i/D+1)]}\right),\ 0 < t < \infty \tag{3.1}$$

Since the class of the best 1:1 codes contains the class of uniquely decodable codes, therefore it follows that

$$L_{1:1}(t) \le L_{UD}(t) \tag{3.2}$$

It may be seen that (3.2) also holds in view of $L_{1:1} \le L_{UD}$.

Now we obtain lower bonds of (3.1) in the next theorem in terms of the Renyi entropy of order $\alpha$ by using the functions defined in theorem 1.

Theorem 2. The exponentiated mean codeword length (3.1) for the best 1:1 code of size $D$ alphabet satisfies the followings:

(a)

$$L_{1:1}(t) \ge H_\alpha(X) - a(1 + \log(H(X) + 1)) - r, \tag{3.3}$$

where $a > \max(1, \log_e D)$, $\tau \ge \log(D^a - 1)/(D^a - D)$, $\alpha = 1/1 + t$ and $0 < t \le 1/a$.

(b)

$$L_{1:1}(t) \ge H_\alpha(X) - a\log(H(X) + 2), \tag{3.4}$$

where $a \ge \max(2, \log_e D)$, $\alpha = 1/1 + t$ and $0 < t \le 1/a$.

(c)

$$L_{1:1}(t) \ge H_\alpha(X) - \log(H(X) + 1)) - \log\log(H(X) + 1) + \ldots\ldots - 6, \tag{3.5}$$

where $\alpha = 1/1 + t$, $0 < t \le 1$ and base of logarithm is 2.

**Proof:**

(a) From (1.4) and (1.5), we have

$(1/t)\log(\sum p_i D^{tl_i}) \ge H_\alpha(X)$

or

$\log(ED^{tl_i}) \ge tH_\alpha(X)$, since $t > 0$.

It imples

$$ED^{tl_i} \ge D^{tH\alpha} \tag{3.6}$$

On using theorem $1(I)$ in (3.6), we get

$E\{D^{t(1+a[\log 1]+\tau)}\} \ge D^t(H\alpha - \tau)$, where $\tau \ge \log(D^a - 1)/(D^a - D)$ and $a > 1$.

or

$E\{D^{tl}\}E\{D^{at(\log l+1)}\} \ge D^{t(H\alpha - \tau)}$, since

$E\{D^{t(1+a(\log l+1))}\} \le E\{D^{tl}\}E\{D^{at(\log l+1)}\}$, or

$E\{D^{tl}\}E\{l^{at}\} \geq \{D^{t(H\alpha-\tau-a)}$

By Jensen's inequality $E\{l^{at}\} \leq (El)^{at}$, so we have

$E\{D^{tl}\}(El)^{at} \geq \{D^{t(H\alpha-\tau-a)}$

or

$$(D^{L1:1(t)})^{t} \geq D^{t(H\alpha-\tau-a)}/(E1)^{at}, \text{ since } E\{D^{tl}\} = (D^{L1:1(t)})^{t} \quad (3.7)$$

Raising both sides of (3.9) to the power 1/t and taking logarithm, we get

$L_{1:1}(t) \geq H_{\alpha}-\tau-a-a\log(E1)$.

Since $L_{1:1} \leq L_{UD} < H(X) + 1$, therefore it follows that

$L_{1:1}(t) \geq H_{\alpha}(X)-a(1 + \log(H(X) + 1) - \tau$

where $a > 1$, $\tau \geq \log(D^{a}-1)/(D^{a}-D)$, $\alpha = 1/1 + t$ and $0 < t \leq 1/a$

(b) From (3.6) and theorem 1 (ii), we have

$E\{D^{t(1+a\log(1+1))}\} \geq tH_{\alpha}(X)$, $a \geq 2$

or

$E\{D^{tl}\}E\{D^{at\log(1+1))}\} \leq D^{tH\alpha}$, since

$E\{D^{t(1+a\log(1+1)}\} \geq E\{D^{tl}\}E\{D^{a+\log(l+1)}\}$

or

$$\{D^{L1:1(t)}\}^{t}E\{(1+1)^{at}\} \geq D^{tH\alpha}, \text{ since } E\{D^{tl}\} \geq (D^{L1:1(t)})^{t} \quad (3.8)$$

By Jensen's inequality $E\{(1+1)^{at}\} \leq E\{1+1)\}^{at}$, so we have

$$(D^{L1:1(t)})^{t} \geq D^{tH\alpha}/(E\{1+1\})^{at} \quad (3.9)$$

Raising both sides of (3.9) to the power 1/t and taking logarithm, we get

$L_{1:1}(t) \geq H_{\alpha}-a\log(E1 + 1)$

It implies.

$L_{1:1}(t) \geq H_{\alpha}(X)-a\log(H)(X) + 2)$, since $El < H(X) + 1$,

where $a \geq 2$, $\alpha = 1/1 + t$ and $0 < t \leq 1/a$

(c) Again from (3.6) and theorem 1(iii), we have

$E\{D^{t(1+\log 1+\log(\log 1)+\ldots+\log(\log(\ldots(\log 1)))+4)}\} \geq D^{tH\alpha}$.

It implies

$E\{D^{tl}\}E(D^{t\log(1+\log 1+\ldots+\log(\ldots(\log 1)))}\} \geq D^{t(H\alpha-4)}$

or

$$(D^{l1:1(t)})^{t}E\{(1^{*})^{t}\} \geq D^{t(H\alpha-4)}, \text{ since } E\{D^{tl}\} = (D^{L1:1(t)})t \quad (3.10)$$

where $1^{*} = 1 \log 1 \log(\log 1)\ldots..\log(\ldots.(\log 1))$.

By Jensen's inequality $E\{(1^{*})^{t}\} \leq (E\{1^{*}\})^{t}$, so we have

$(D^{L1:1(t)})t \geq D^{t(H\alpha-4)}/(E\{1^{*}\})^{t}$

Raising both sides of (3.11) to the power of 1/t and taking logarith, we have

$L_{1:1}(t) \geq H_\alpha$–4–log $E\{1^*\}$

or

$$L_{1:1}(t) \geq H_\alpha–4–E\{\log(1^*)\}, \text{ since } E\{\log(1^*)\} \geq \log E\{1^*\} \quad (3.12)$$

we consider

$\log(1^*) = \log 1 + \log(\log 1) + ....$ upto the last positive term = $\log^* 1$ (say Although $\log^* 1$ is not concave, yet Leung-Yan-Cheang and Cover [4] proved that there exists a concave function $F^*(1)$ such that

$F^*(1) \leq \log^* 1 < F^*(1) + 2$.

Thus

$$E\{\log(1^*)\} = E\{\log^* 1\} \leq E\{F^*(1) + 2\} \leq F^*(E1) + 2 \leq \log^*(E1) + 2 \quad (3.13)$$

Substituting (3.13) in (3.12), we get

$L_{1:1}(t) \geq H_\alpha(X)–6–\log^*(E1)$

or

$L_{1:1}(t) \geq H_\alpha(X)–6–\log(E1)–\log\log(E1).....$

Since $E1 < H(X) + 1$, therefore it follows that

$L_{1:1}(t) \geq H\alpha\,(X)–6–\log(H(X) + 1)–\log\log(H(X) + 1)...$

It may be noted that part (c) has been proved by taking arbitrary base D of logarithm. Thus it holds for D = 2 also. This completes the proof of theorem 2.

Particular case: It can be easily verified that (3.3), (3.4) and (3.5) reduce to the results due to Leung-Yan-Cheong and Cover [4] for Shannon entropy, when $\alpha \to 1$ and $D = 2$.

From (3.2) and (1.5) it follows that

$$L_{1:1}(t) < H_\alpha(X) + 1 \quad (3.16)$$

Hence (3.16) gives an upper bound on $L_{1:1}(t)$.

## Remarks

The upper bound on $L_{1:1}(t)$ is equal to that of $L_{UD}(t)$ while the lower bounds are better than lower bound on $L_{UD}(t)$. The lower bounds obtained in this paper are more general due to a $\alpha$ parameter and thus are more effective and flexible for application point of view.

## Acknowledgement

The author expresses his sincere thanks to Professor W. Oettli, Universitat Mannheim, for his valuable discussions during preparation of this paper.

## References

1. Campbell, L. L. A coding theorem and Renyi's entropy. Information and Control 8 (1965), 423–429.

2. JELINEK, F. Buffer overflow in variable length coding of fixed rate sources. IEEE Vol. I T 14 (1980), 490–501.

3. KIEFFER, J. C. Variable length source coding with a cost depending only on the codeword length. Information and Control 41 (1979), 136–146.

4. LEUNG-YAN-CHEONG, S. K., COVER, T. M. Some equivalence between Shannon entropy and Kolmogorov Complexity. IEEE Vol. I T 24 (1978), 331–338.

5. RENYI, A. On measures of entropy and information. Proceeding 4th Berk. Symp. Math. Stat. Prob., Vol. 1(1961), 546–561.