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

D. S. HODDA
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.


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

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

where $H(X)$ is the Shannon entropy of the random variable $X$ and $L_{UD} = \sum p_i \cdot i$ 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

$$L_{1:1} = \sum p_i [\log(i/2 + 1)] \geq H(X) - \log \log N - 3,$$  \hspace{1cm} (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.
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.

Campbell [1] introduced the exponentiated codeword length.

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

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:

$$H_a(X) \leq L_{UD}(t) < H_a(X) + 1$$

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

where $H_a(X)$ 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 $H_a(X)$ 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 codeword 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 $P = \{(p_1, p_2, ..., p_N)\}$, $p_i \geq 0$ for all $i$, $\Sigma p_i = 1$, $p_i \geq p_j$ for $i < j$. Let $l_i, i = 1, 2, ..., N$, 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

$$l_1 = 1, l_2 = 1, ..., l_D = 1, l_{D+1} = 2, ..., l_{D(D+1)} = 2, ..., etc$$
Thus by inspection we can see that

\[ l_i = \lfloor \log(D-1)i/D + 1 \rfloor \]  

(2.1)

where \( \lfloor S \rfloor \) denotes the smallest integer greater than or equal to \( S \).

Now we define a function \( h(l_i) \) such that \( \sum D^{-h(l_i)} \leq 1 \) 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 \( \sum_{i=1}^{N} D^{-h(l_i)} > 1 \), 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) \( h(l_i) = l_i + a[\log(l_i)] + \log(D^u - 1)/ (D^u - D) \), \( a > 1, D \geq 2 \).  

(2.2)

(ii) \( h(l_i) = l_i + a[\log(l_i + 1)] \), \( a > 2 \) and \( D \geq 2 \).  

(2.3)

(iii) \( h(l_i) = l_i + [\log(l_i + \log(l_i) + \ldots + \log(\log(\ldots(\log(l_i))))\ldots)] + 4 \).  

where we only take the first \( K \) iterates for which

\[ \log(\log(\ldots(\log(l_i)))) \] is positive.

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

Lemma 2.1 Let

\[ G_D(X) = 1 \times \log_D \times \log_D(\log_D X) \]  

(2.8)

then

\[ I_D = G_D(x) = \int_1^a \frac{dx}{x \log_D X \log(\log_D X) \ldots} \begin{cases} \text{infinite if } D \geq e \\ \text{finite if } D < e \end{cases} \]

For proof refer to [4]

Thus \( I_D \) is finite only if \( D < e \) which implies \( D = 2 \).

For \( D = 2 \) and \( M = \Log_2 e \), we have.

\[ I_2 \leq \log_2 e / (\log_2 e - 1) < 3.26 \]  

(2.9)

Hence

\[ \sum 1/\log \log(\log 1) \ldots < I_2 + 1 < 5. \]  

(2.10)

Substituting (2.10) in (2.7), we get

\[ S < 5.2^{-c+1} \]  

(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

\[
L_{1:1}(t) = \frac{1}{t} \log \left( \sum p_i D_i^{\alpha(\log(D-1)/(D+1))} \right), \quad 0 < t < \infty
\]  
(3.1)

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

\[
L_{1:1}(t) \leq L_{UD}(t)
\]  
(3.2)

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

Now we obtain lower bounds 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) 
\[
L_{1:1}(t) \geq H_\alpha(X) - \alpha(1 + \log (H(X) + 1)) - t,
\]  
(3.3)

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

(b) 
\[
L_{1:1}(t) \geq H_\alpha(X) - \log (H(X) + 2),
\]  
(3.4)

where \( \alpha \geq \max(2, \log, D) \), \( \alpha = 1/1 + t \) and \( 0 < t \leq 1/\alpha \).

(c) 
\[
L_{1:1}(t) \geq H_\alpha(X) - \log (H(X) + 1) - \log \log(H(X) + 1) + \ldots - 6,
\]  
(3.5)

where \( \alpha = 1/1 + t \), \( 0 < t \leq 1 \) and base of logarithm is 2.

Proof:

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

\((1/t)\log(\sum p_i D_i^{\alpha}) \geq H_\alpha(X)\)

or

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

It implies

\(ED^{\alpha} \geq D^{tH_\alpha}\)  
(3.6)

On using theorem 1(f) in (3.6), we get

\(E[D^{(1-a(\log 1+1))}] \geq D^\tau(H_\alpha - \tau)\), where \( \tau \geq \log(D^{\alpha-1}/(D^\alpha-D)) \) and \( a > 1 \).

or

\(E[D^\alpha]E[D^a(\log 1+1)] \geq D^\tau(H_\alpha - \tau)\), since

\(E[D^{(t-a(\log 1+1))}] \leq E[D^t]E[D^a(\log 1+1)]\).
or
$$E[D^I]E[I^I] \geq |D^{H_0-\tau-a}|$$

By Jensen’s inequality $E[I^I] \leq (EI)^I$, so we have
$$E[D^I](EI)^I \geq |D^{H_0-\tau-a}|$$

or
$$|D^{L_1(t)}|^I \geq D^{H_0-\tau-a}(EI)^I, \text{ since } E[D^I] = (D^{L_1(t)})^I \tag{3.7}$$

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 $L_{1,1} \leq L_{1,0} < H(X) + 1$, therefore it follows that
$$L_{1,1}(t) \geq H_{\alpha}(X) - \alpha(1 + \log(H(X) + 1) - \tau$$

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

(b) From (3.6) and theorem 1 (ii), we have
$$E[D^I]E[I^I] \geq tH_a(X), a \geq 2$$

or
$$E[D^I]E[D^{a \log(1 + 1)}] \leq D^{H_a}, \text{ since }$$
$$E[D^I]E[D^{a \log(1 + 1)}] \geq E[D^I]E[D^a \log(t - 1)]$$

or
$$|D^{L_1(t)}|^I \geq D^{H_a}(E[1 + 1]^I), \text{ since } E[D^I] \geq (D^L(1))^I \tag{3.8}$$

By Jensen’s inequality $E[(1 + 1)^I] \leq E[1 + 1]^I$, so we have
$$|D^{L_1(t)}|^I \geq D^{H_a}(E[1 + 1]^I). \tag{3.9}$$

Raising both sides of (3.9) to the power $1/t$ and taking logarithm, we get
$$L_{1,1}(t) \geq H_{a} - a \log(E1 + 1)$$

It implies.
$$L_{1,1} \geq H_{a}(X) - \alpha \log(H(X) + 2), \text{ since } E1 < H(X) + 1,$$

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
$$E[D^I]E[D^{1 \cdots \log(1 + 1) \cdots \log(1 + 1)}] \geq D^{H_a}.$$ 

It implies
$$E[D^I]E[D^{1 \cdots \log(1 + 1) \cdots \log(1 + 1)}] \geq D^{H_a-4}$$

or
$$|D^{L_1(t)}|^I \geq D^{H_a-4}, \text{ since } E[D^I] = (D^{L_1(t)})^I \tag{3.10}$$

where $1^* = 1 \log (1) \log(\log (1)) \cdots \log(\log(\log(1)))$.

By Jensen’s inequality $E[1^*]^I \leq (E[1^*])^I$, so we have
$$|D^{L_1(t)}|^I \geq D^{H_a-4}(E[1^*])^I$$

Raising both sides of (3.11) to the power $1/t$ and taking logarithm, we have
\(L_{1;1}(t) \geq H_\alpha - 4 - \log E[1^\phi]\)

or

\[L_{1;1}(t) \geq H_\alpha - 4 - E[\log(1^\phi)], \text{ since } E[\log(1^\phi)] \geq \log E[1^\phi]\] (3.12)

we consider

\(\log(1^\phi) = \log 1 + \log(\log 1) + \ldots \text{ upto the last positive term } = \log^\phi 1 \) (say Although \(\log^\phi 1\) is not concave, yet Leung-Yan-Cheong and Cover [4] proved that there exists a concave function \(F^\phi(1)\) such that

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

Thus

\[E[\log(1^\phi)] = E[\log^\phi 1] \leq E[F^\phi(1) + 2] \leq F^\phi(E1) + 2 \leq \log^\phi(E1) + 2 \] (3.13)

Substituting (3.13) in (3.12), we get

\[L_{1;1}(t) \geq H_\alpha(X) - 6 - \log^\phi(E1)\]

or

\[L_{1;1}(t) \geq H_\alpha(X) - 6 - \log(E1) - \log(\log(E1))\ldots\]

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

\[L_{1;1}(t) \geq H_\alpha(X) - 6 - \log(H(X) + 1) - \log(\log(H(X) + 1))\ldots\]

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

\[L_{1;1}(t) < H_\alpha(X) + 1\] (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 \(L_{UP}(t)\). 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

2. Jelinek, F.

3. Kieffer, J. C.

4. Leung-Yan-Cheong, S. K.
   Cover, T. M.

5. Rynne, A.


Variable length source coding with a cost depending only on the codeword length. Information and Control 41 (1979), 136–140.
