New Bounds on the Expected Length of One-to-One Codes *

Carlo Blundo and Roberto De Prisco

Dipartimento di Informatica ed Applicazioni
Università di Salerno, 84081 Baronissi (SA), Italy.
{carblu,robdep}@dia.unisa.it

Abstract

In this correspondence we provide new bounds on the expected length L of a binary one-to-one code for a discrete random variable X with entropy H. We prove that $L \geq H - \log(H+1) - H\log(1+1/H)$. This bound improves on previous results. Furthermore, we provide upper bounds on the expected length of the best code as function of H and the most likely source letter probability.

Index Terms – Source coding, one-to-one codes, non-prefix codes.

1 Introduction

Let X be a discrete random variable which assumes values on a countable support set $\mathcal{X}$. A binary encoding for X is a function that maps each element of $\mathcal{X}$ to a binary codeword. Without loss of generality, assume that $\mathcal{X} = \{1, 2, ..., N\}$, where N can be infinite. The probability that X takes the value i is p_i. Throughout this paper we assume, without loss of generality, that $p_i \geq p_{i+1}$. Given an encoding, the expected length of the encoding is

$$L(X) = \sum_{i=1}^{N} p_i n_i \tag{1}$$

where n_i is the length of the codeword used to encode the value $i \in \mathcal{X}$. The entropy of X is denoted $H(X)$ or simply H when X is clear from the context. Shannon [3] proved that the minimum expected length L_{pre} of a prefix-free encoding of X satisfies $H \leq L_{pre} \leq H + 1$.

*This work is partially supported by Italian Ministry of University and Research (M.U.R.S.T.) and by National Council for Research (C.N.R.).

Prefix-free codes are very useful as they are instantaneously decodable. Relaxing the prefix property and keeping the non-singularity of the code (i.e., the encoding is a one-to-one mapping), we obtain a larger class of possible encodings. Clearly, the expected length of the best one-to-one encoding is at most L_{pre}.

Two different frameworks have been considered in the literature depending on whether the empty codeword is used. We refer to encodings using the empty codeword as $\{0,1\}^*$-encodings and to encodings not using the empty codeword as $\{0,1\}^+$-encodings.

$\{0,1\}^*$-encodings. If ϵ is used then the codebook is $\{\epsilon, 0, 1, 00, 01, 10, 11, 000, ...\}$. Clearly, the best $\{0,1\}^*$-encoding uses the N shortest codewords of the codebook. It is easy to see that the length of the i-th shortest codeword is

$$n_i = \lfloor \log i \rfloor. \tag{2}$$

All logarithms in this paper are to the base 2. We denote by L_ϵ the expected length of the best $\{0,1\}^*$-encoding.

Recently, Alon and Orlitsky [1] proved that

$$L_\epsilon \geq H - \log(H+1) - \log e. \tag{3}$$

Wyner [5] proved that

$$L_\epsilon \leq H. \tag{4}$$

Bound (4) is achieved by the constant random variable.

In this correspondence we prove that

$$H \leq L_\epsilon + (L_\epsilon + 1)\mathcal{H}\left(\frac{1}{L_\epsilon + 1}\right) \tag{5}$$

extending the result of [4]. Previous bound is achievable for any value of H.

We also provide an explicit lower bound on L_ϵ,

$$L_\epsilon \geq H - \log(H+1) - H \log\left(1 + \frac{1}{H}\right). \tag{6}$$

improving on (3).

We denote by $\mathcal{H}$ the binary entropy. We prove that

$$L_\epsilon \leq \begin{cases} H + p_1 \log p_1 & \text{for } 0 < p_1 \leq 0.5 \\ H + 1 - p_1 - \mathcal{H}(p_1) & \text{for } 0.5 < p_1 \leq 1 \end{cases} \tag{7}$$

slightly improving on (4).

$\{0,1\}^+$-encodings. If ϵ is not used then the codebook, that is the set of possible codewords, is $\{0, 1, 00, 01, 10, 11, 000, 001, ...\}$. Clearly, the best $\{0,1\}^+$-encoding uses the N shortest codewords of the codebook. It is easy to see that the length of the i-th shortest codeword is

$$n_i = \left\lceil \log \left(\frac{i}{2} + 1 \right) \right\rceil . \tag{8}$$

We denote by L the expected length of the best $\{0,1\}^+$-encoding.

Leung-Yan-Cheong and Cover [2] proved the following bound

$$L \geq H - \log^*(H+1) - 6 \tag{9}$$

where $\log^* x \triangleq \log x + \log\log x + \cdots$ stopping at last positive term. Alon and Orlitsky's bound (3) improves on previous bound (any lower bound for L_ϵ holds for L as well).

Verriest [4] proved that when $H \geq 1$

$$H \leq L \left(1 + \mathcal{H} \left(\frac{1}{L} \right) \right) . \tag{10}$$

Since $L \leq L_{pre}$ we have that

$$L \leq H + 1. \tag{11}$$

The above bound is achieved by the constant random variable.
Since $L \geq L_\epsilon$, bound (6) holds for L, too. Thus, we have

$$L \geq H - \log(H+1) - H \log \left(1 + \frac{1}{H} \right) . \tag{12}$$

Moreover, improving on (11), we prove that

$$L \leq \begin{cases} H + \frac{\ell p_1}{2} + \frac{1+p_1}{2^{\ell+1}} + p_1 \log p_1 & \text{for } 0 \leq p_1 \leq 0.5 \\ H + \frac{5-p_1}{4} - \mathcal{H}(p_1) & \text{for } 0.5 < p_1 \leq 1, \end{cases} \tag{13}$$

where $\ell = \lfloor \log(1 + 1/p_1) \rfloor$.

This correspondence is organized as follows. In Section 2 we consider $\{0,1\}^*$-encodings and we prove the upper and the lower bounds on L_ϵ. In Section 3 we address the case of $\{0,1\}^+$-encodings providing the upper and the lower bounds on L.

2 $\{0,1\}^*$-encodings

In this section we consider the case of encodings that do use the empty codeword. We employ the technique of [4] based on Lagrange's multipliers. We will prove that L_ϵ is related to the entropy H by

$$H \leq L_\epsilon + (L_\epsilon + 1) \mathcal{H} \left(\frac{1}{L_\epsilon + 1} \right) . \tag{14}$$

Before showing how to obtain (14) we need some technical lemmas. In the next lemma, given a random variable X of fixed entropy H taking on infinitely many values, we compute the probability distribution of entropy H minimizing (1) (i.e., among all probability distributions of a given fixed entropy H, we want to find the one minimizing the expected length).

Lemma 2.1 *Let X be a discrete random variable with entropy H. Suppose that X takes on infinitely many values. The function $L(X) = \sum_{i=1}^{\infty} p_i n_i$, where $n_i = \lfloor \log i \rfloor$, achieves the minimum at*

$$p_i = p_1 \left(\frac{1-p_1}{2}\right)^{n_i}.$$

Proof. We want to minimize the function $L(X) = \sum_{i=1}^{\infty} p_i n_i$ under the constraints

$$\sum_{i=1}^{\infty} p_i = 1 \tag{15}$$

and

$$\sum_{i=1}^{\infty} p_i \log \frac{1}{p_i} = H. \tag{16}$$

Using the Lagrange's multipliers λ and μ, the Lagrangian $\mathcal{L}$ to be studied is the following.

$$\mathcal{L} = \sum_{i=1}^{\infty} p_i n_i + \lambda \sum_{i=1}^{\infty} p_i - \mu \sum_{i=1}^{\infty} p_i \log p_i.$$

For the optimality we get

$$\frac{\partial \mathcal{L}}{\partial p_i} = n_i + \lambda - \mu(\log p_i + \log e) = 0.$$

Therefore, $p_i = 2^{n_i/\mu + (\lambda - \mu \log e)/\mu}$. Setting $\alpha = 2^{(\lambda - \mu \log e)/\mu}$ and $\beta = 1/\mu$ we obtain that $p_i = \alpha 2^{\beta n_i}$. Recalling that $n_1 = \lfloor \log 1 \rfloor = 0$ we get that $\alpha = p_1$. Hence, $p_i = p_1 2^{\beta n_i}$. For $j \geq 0$, with N_j we denote the number of codewords of length j. Therefore, $N_j = 2^j$. Substituting the p_i's in (15), we find the value of β. Indeed,

$$\begin{aligned} 1 &= \sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} p_1 2^{\beta n_i} = p_1 \sum_{j=0}^{\infty} N_j 2^{\beta j} = p_1 (1 + \sum_{j=1}^{\infty} N_j 2^{\beta j}) \\ &= p_1 \left(1 + \sum_{j=1}^{\infty} (2^{\beta+1})^j\right) = p_1 \left(1 + \frac{2^{\beta+1}}{1 - 2^{\beta+1}}\right) = \frac{p_1}{1 - 2^{\beta+1}}. \end{aligned}$$

Hence, $p_1 = 1 - 2^{\beta+1}$ from which we obtain that $\beta = \log \frac{1-p_1}{2}$. Therefore, $p_i = p_1 \left(\frac{1-p_1}{2}\right)^{n_i}$ and the lemma is proved. □

In the following we will show how to express the entropy H of a random variable with probability distribution $p_i = p_1(\frac{1-p_1}{2})^{n_i}$, in terms of $x = 1 - p_1$ and $\mathcal{H}(x)$. If we substitute the values of the p_i's we get

$$H + \log p_1 + \log \frac{1-p_1}{2} \sum_{i=1}^{\infty} n_i p_1 \left(\frac{1-p_1}{2}\right)^{n_i} = 0. \tag{17}$$

Considering the expected length $\hat{L}_\epsilon$ of the p_i's, we have

$$\begin{aligned} \hat{L}_\epsilon &= \sum_{i=1}^{\infty} n_i p_1 \left(\frac{1-p_1}{2}\right)^{n_i} = \sum_{j=0}^{\infty} j N_j p_1 \left(\frac{1-p_1}{2}\right)^{j} \\ &= \sum_{j=0}^{\infty} j 2^j p_1 \left(\frac{1-p_1}{2}\right)^{j} = p_1 \sum_{j=0}^{\infty} j(1-p_1)^j = \frac{1-p_1}{p_1}. \end{aligned} \tag{18}$$

Substituting (18) in (17) after some algebra we get

$$H + \frac{p_1 \log p_1 + (1-p_1)\log(1-p_1) - (1-p_1)}{p_1} = 0. \tag{19}$$

Setting $x = 1 - p_1$ we obtain

$$H = \frac{\mathcal{H}(x) + x}{1-x}. \tag{20}$$

In [4] it has been proved that the function $y(x) = \frac{\mathcal{H}(x)+x}{1-x}$, defined in the interval $[0,1[$ is invertible. It follows that for any given value of the entropy H there exists an unique solution in $[0,1[$ of the equation (20).

Next we prove

Theorem 2.2 *For any discrete random variable X, the entropy H and the expected length L_ϵ of the best $\{0,1\}^*$-encoding, are related by*

$$H \leq L_\epsilon + (L_\epsilon + 1)\mathcal{H}\left(\frac{1}{L_\epsilon + 1}\right).$$

Proof. Suppose that the random variable X takes on infinitely many values and that its entropy is H. Let $\hat{L}_\epsilon$ be the achievable minimum average codeword length, that is, the minimum of (1) seen as function of the p_i's once that the value assumed by the entropy H has been fixed. The achievable minimum average codeword length $\hat{L}_\epsilon$ has been computed in (18). Recalling that $x = 1 - p_1$, we obtain

$$\hat{L}_\epsilon = \frac{x}{1-x} \quad \text{or equivalently} \quad x = 1 - \frac{1}{\hat{L}_\epsilon + 1}.$$

Since $\mathcal{H}(x) = \mathcal{H}(1-x)$, substituting $x = 1 - 1/(\hat{L}_\epsilon + 1)$ in (20), we get

$$H = \hat{L}_\epsilon + (\hat{L}_\epsilon + 1)\mathcal{H}\left(\frac{1}{\hat{L}_\epsilon + 1}\right). \tag{21}$$

Let Y be a random variable having the same entropy as X. Clearly, the expected length L_ϵ of the best $\{0,1\}^*$-encoding of Y is at least $\hat{L}_\epsilon$. Since the function $x+(x+1)\mathcal{H}(1/(x+1))$ is a non-decreasing function, we get

$$H \leq L_\epsilon + (L_\epsilon+1)\mathcal{H}\left(\frac{1}{L_\epsilon+1}\right)$$

and the theorem is proved. □

The next theorem gives a bound on the expected length L_ϵ of any $\{0,1\}^*$-encoding of a random variable X in term of the entropy H of X.

Theorem 2.3 *For any discrete random variable X, with entropy H, the expected length L_ϵ of the best $\{0,1\}^*$-encoding satisfies*

$$L \geq H - \log(H+1) - H\log(1+\frac{1}{H}).$$

Proof. We get

$$\begin{aligned}
L_\epsilon &\geq H-(L_\epsilon+1)\mathcal{H}\left(\frac{1}{L_\epsilon+1}\right) \text{ (from Theorem 2.2)}\\
&\geq H-(H+1)\mathcal{H}\left(\frac{1}{H+1}\right) \text{ (from (4))}\\
&= H-\log(H+1)-H\log\left(1+\frac{1}{H}\right).
\end{aligned}$$

□

A simple algebra shows that the above bound improves on (3) for any value of H (though the difference between the two bounds tends to 0 as H tends to infinity).

In the following we will prove an upper bound on L_ϵ. The only known upper bound is the Wyner's upper bound (4).

Theorem 2.4 *For any discrete random variable X, with entropy H, the expected length L_ϵ of the best $\{0,1\}^*$-encoding satisfies*

$$L_\epsilon \leq \begin{cases} H - p_1\log\frac{1}{p_1} & \text{for } 0 < p_1 \leq 0.5 \\ H+1-p_1-\mathcal{H}(p_1) & \text{for } 0.5 < p_1 \leq 1. \end{cases}$$

Proof. First, we prove that

$$L_\epsilon \leq \sum_{i\geq 2} p_i \log i - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i. \tag{22}$$

Indeed, recalling that $n_1 = \lfloor \log 1 \rfloor = 0$, we have

$$\begin{aligned} L_\epsilon &= \sum_{i\geq 1} p_i n_i = \sum_{i\geq 1} p_i \lfloor \log i \rfloor = \sum_{\substack{i\neq 2^j-1\\ j\geq 1}} p_i \lfloor \log i \rfloor + \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \lfloor \log i \rfloor \\ &= \sum_{\substack{i\neq 2^j-1\\ j\geq 1}} p_i \lfloor \log i \rfloor + \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \log i - \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i (\log i - \lfloor \log i \rfloor) \\ &\leq \sum_{i\geq 2} p_i \log i - \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i (\log i - \lfloor \log i \rfloor) \end{aligned}$$

The function $\log(2^j-1) - \lfloor \log(2^j-1) \rfloor$ is an increasing function of j. For $j \geq 2$, it reaches its minimum at $j = 2$. This minimum is equal to $\log 3 - \lfloor \log 3 \rfloor > 0.5$. Thus, relation (22) holds.

Next, we prove that

$$L_\epsilon \leq H - p_1 \log \frac{1}{p_1} - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i. \tag{23}$$

Indeed, since $p_i \leq 1/i$, we have

$$\begin{aligned} L_\epsilon &\leq \sum_{i\geq 2} p_i \log i - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \\ &\leq \sum_{i\geq 2} p_i \log \frac{1}{p_i} - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \\ &= H - p_1 \log \frac{1}{p_1} - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i. \end{aligned}$$

Finally, we prove that

$$L_\epsilon \leq H - \mathcal{H}(p_1) + 1 - p_1 - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i. \tag{24}$$

Indeed, observing that for any $i \geq 2$, it holds that $p_i \leq (1-p_1)/(i-1) \leq 2(1-p_1)/i$, we get

$$\begin{aligned} L_\epsilon &\leq \sum_{i\geq 2} p_i \log i - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \leq \sum_{i\geq 2} p_i \log \frac{2(1-p_1)}{p_i} - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \\ &= \sum_{i\geq 2} p_i \log \frac{1}{p_i} + (1 + \log(1-p_1)) \sum_{i\geq 2} p_i - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \\ &= H - p_1 \log \frac{1}{p_1} + (1 + \log(1-p_1))(1-p_1) - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i \\ &= H - \mathcal{H}(p_1) + 1 - p_1 - 0.5 \sum_{\substack{i=2^j-1\\ j\geq 2}} p_i. \end{aligned}$$

From (23) and (24), since $\sum_{\substack{i=2^j-1 \\ j\geq 2}} p_i \geq 0$, we get

$$L_\epsilon \leq H - p_1 \log \frac{1}{p_1}$$

and

$$L_\epsilon \leq H - \mathcal{H}(p_1) + 1 - p_1.$$

It is easy to check that $p_1 \log \frac{1}{p_1} \geq \mathcal{H}(p_1) + p_1 - 1$ for $0 < p_1 \leq 0.5$. Thus, the theorem holds. $\square$

The above bound improves on (4) for $0 < p_1 < 1$; when $p_1 = 1$ they coincide.

3 $\{0,1\}^+$-encodings

In this section we consider the case of encodings that do not use the empty codeword. The following relation between L and L_ϵ holds:

$$L = L_\epsilon + \sum_{i\geq 1} p_{2^i-1}. \tag{25}$$

Clearly $L_\epsilon < L \leq L_\epsilon + 1$. Therefore, the bound (6) holds for $\{0,1\}^+$-encodings as well, and from (7) we obtain

$$L \leq \begin{cases} H + 1 + p_1 \log p_1 & \text{for } 0 < p_1 \leq 0.5 \\ H + 2 - p_1 - \mathcal{H}(p_1) & \text{for } 0.5 < p_1 \leq 1 \end{cases} \tag{26}$$

improving on (11).

In the sequel we improve previous upper bound, providing a sharper upper bound on $\sum_{i\geq 1} p_{2^i-1}$.

Lemma 3.1 *For any integer $\ell \geq 1$ it holds that*

$$\sum_{i\geq 1} p_{2^i-1} \leq (\ell - 1)p_1 + \frac{1+p_1}{2^\ell}.$$

Proof. Recall that the p_i's are in non-increasing order. First, we prove that for any $\ell \geq 1$

$$\sum_{i\geq \ell+1} p_{2^i-1} \leq \frac{1 - \sum_{j=1}^{2^\ell-1} p_j}{2^\ell} \tag{27}$$

Indeed,

$$1 - \sum_{j=1}^{2^\ell-1} p_j = \sum_{j\geq 2^\ell} p_j \geq \sum_{i\geq \ell+1} \sum_{k=1}^{2^\ell-1} p_{2^i-1-k} \geq \sum_{i\geq \ell+1} \sum_{k=1}^{2^\ell-1} p_{2^i-1} = 2^\ell \sum_{i\geq \ell+1} p_{2^i-1}.$$

Next, we prove that for any $\ell \geq 1$

$$\sum_{j=1}^{2^\ell-1} p_j \geq \sum_{i=1}^{\ell} 2^{i-1} p_{2^i-1}. \tag{28}$$

In fact,

$$\sum_{j=1}^{2^\ell-1} p_j = \sum_{i=1}^{\ell} \sum_{k=0}^{2^{i-1}-1} p_{2^i-1-k} \geq \sum_{i=1}^{\ell} \sum_{k=0}^{2^{i-1}-1} p_{2^i-1} = \sum_{i=1}^{\ell} 2^{i-1} p_{2^i-1}.$$

Whence,

$$\begin{aligned}
\sum_{i\geq 1} p_{2^i-1} &= \sum_{i=1}^{\ell} p_{2^i-1} + \sum_{i\geq \ell+1} p_{2^i-1} \\
&\leq \sum_{i=1}^{\ell} p_{2^i-1} + \frac{1}{2^\ell} - \sum_{i=1}^{\ell} 2^{i-1-\ell} p_{2^i-1} \quad \text{(from (27) and (28))} \\
&= \frac{1}{2^\ell} + \sum_{i=1}^{\ell} (1 - 2^{i-\ell-1}) p_{2^i-1} \\
&\leq \frac{1}{2^\ell} + p_1 \sum_{i=1}^{\ell} (1 - 2^{i-\ell-1}) \\
&= (\ell-1)p_1 + \frac{1+p_1}{2^\ell}.
\end{aligned}$$

Thus, the lemma holds. □

A simple algebra shows that the bound of Lemma 3.1 for $\ell+1$ is sharper than the one for ℓ if and only if $p_1 \leq \frac{1}{2^{\ell+1}-1}$. This implies that when $p_1 \in]1/(2^{\ell+1}-1), 1/(2^\ell-1)]$ the best bound is the one for ℓ. Therefore, for any p_1 the value of ℓ achieving the sharpest bound is $\ell = \lfloor \log(1+1/p_1) \rfloor$. Thus, we have

$$\sum_{i\geq 1} p_{2^i-1} \leq (\ell-1)p_1 + \frac{1+p_1}{2^\ell}, \tag{29}$$

where $\ell = \lfloor \log(1+1/p_1) \rfloor$.

We can use the bound of Lemma 3.1 to improve on bound (26). Indeed, from (23), (24), and (25) we have that

$$L \leq \begin{cases} H + p_1 \log p_1 + p_1/2 + 0.5 \sum_{i\geq 1} p_{2^i-1} & \text{for } 0 < p_1 \leq 0.5 \\ H + 1 - p_1/2 - \mathcal{H}(p_1) + 0.5 \sum_{i\geq 1} p_{2^i-1} & \text{for } 0.5 < p_1 \leq 1. \end{cases}$$

From (29) we obtain the following theorem.

Theorem 3.2 *For any discrete random variable X with entropy H, the expected length L of the best $\{0,1\}^+$-encoding satisfies*

$$L \leq \begin{cases} H + \frac{\ell p_1}{2} + \frac{1+p_1}{2^{\ell+1}} + p_1 \log p_1 & \textit{for } 0 < p_1 \leq 0.5 \\ H + \frac{5-p_1}{4} - \mathcal{H}(p_1) & \textit{for } 0.5 < p_1 \leq 1, \end{cases}$$

where $\ell = \lfloor \log(1+1/p_1) \rfloor$.

Previous bound improves on (26). Figure 1 shows upper bounds on $L - H$ as provided by (13) and (26).

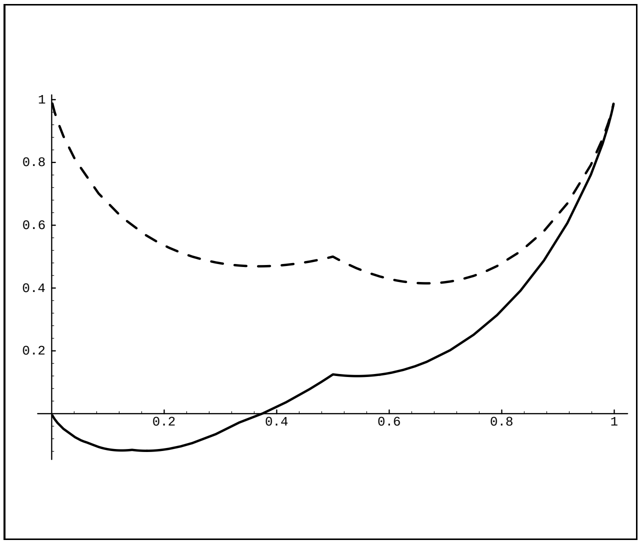

Figure 1: Upper bounds on $L - H$ as provided by (26), dashed, and (13), solid.

Finally, we remark that bound (22) can be further improved either by simply using a more accurate lower bound on the function $\log(2^j - 1) - \lfloor \log(2^j - 1) \rfloor$ (for example 0.58496 instead of 0.5), or by providing a lower bound on $\log(2^j - 1) - \lfloor \log(2^j - 1) \rfloor$ for $j \geq k$ and considering the first k terms apart. As an example, fixed $k = 3$, for any $j \geq 3$, we have that $\log(2^j - 1) - \lfloor \log(2^j - 1) \rfloor \geq \log 7 - \lfloor \log 7 \rfloor > 0.80735$). However, the improved bounds that can be obtained slightly differ from the ones provided in Theorem 3.2 and their expressions are quite complicate.

Acknowledgements

The authors would like to express their gratitude to Alfredo De Santis and Ugo Vaccaro for helpful discussions and suggestions. The authors wish to thank Alon Orlitsky and one of the referees for useful comments on this correspondence. Alon Orlitsky pointed out that the bound of Theorem 2.2 can be obtained using the derivation of Lemma 1 in [1], noticing that $H = \log E + (E-1)\log(E/(E-1))$, where $E = E(X)$, and continuing as in that proof.

References

[1] N. Alon and A. Orlitsky, "A Lower Bound on the Expected Length of One-to-One Codes", *IEEE Trans. Inform. Theory*, vol. IT–40, no. 5, pp. 1670–1672, Sept. 1994.

[2] S. K. Leung-Yan-Cheong and T. M. Cover, "Some Equivalences Between Shannon Entropy and Kolmogorov Complexity", *IEEE Trans. Inform. Theory*, vol. IT–24, no. 3, pp. 331–338, May 1978.

[3] C. Shannon, "A Mathematical Theory of Communication", *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948

[4] E. I. Verriest, "An Achievable Bound for Optimal Noiseless Coding of a Random Variable", *IEEE Trans. Inform. Theory*, vol. IT–32, no. 4, pp. 592–594, July 1986.

[5] A. D. Wyner, "An Upper Bound on the Entropy Series", *Inform. Control*, vol. 20, pp. 176–181, 1972.