A Lower Bound on the Expected Length of One-to-One Codes

Noga Alon* Alon Orlitsky†

August 10, 1994

Abstract

We show that the expected length of any one-to-one encoding of a discrete random variable X is at least[1] $H(X)-\log(H(X)+1)-\log e$ and that this bound is asymptotically achievable.

Key words: Source coding, one-to-one codes, non-prefix codes

1 Introduction

Let X be a random variable distributed over a countable support set $\mathcal{X}$. A *(binary, one-to-one) encoding* of X is an injection $\phi : \mathcal{X} \to \{0,1\}^*$, the set of finite binary strings. The expected number of bits ϕ uses to encode X is

$$l(\phi) \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} \Pr(x)|\phi(x)|$$

where $\Pr(x)$ is the probability that $X = x$ and $|\phi(x)|$ is the length of $\phi(x)$.

A string $x_1,\ldots,x_m$ is a *prefix* of a string $y_1,\ldots,y_n$ if $m \le n$ and $x_i = y_i$ for $i = 1,\ldots,m$. Usually, one is interested in *prefix-free encodings* where no string in $\phi(\mathcal{X})$ is a prefix of another. Let

$$L(X) \stackrel{\text{def}}{=} \min\{l(\phi) \ : \ \phi \text{ is a prefix-free encoding of } X\}$$

*AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974 and Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel.

†AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974.

[1]Throughout, logarithms are to the base 2 and $H(X) \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} \Pr(x)\log\frac{1}{\Pr(x)}$ is the binary entropy of X.

denote the minimum expected number of bits used in a prefix-free encoding of X. Shannon [1] showed[2] that for all discrete random variables X,

$$H(X) \leq L(X) \leq H(X) + 1 \ .$$

Occasionally, encodings that are not necessarily prefix free are encountered. This is the case, for example, if there is an "end of message" symbol. It is therefore of interest to determine

$$\ell(X) \stackrel{\text{def}}{=} \min\{l(\phi) \ : \ \phi \text{ is an encoding of } X\},$$

the minimum expected number of bits used in any one-to-one encoding of X.

Wyner [2] proved that for all discrete random variables X,

$$\ell(X) \leq H(X).$$

This bound, named *Wyner's upper bound* by Elias [3], is achieved by the constant random variables. Leung-Yan-Cheong and Cover [4] proved that for all discrete random variables X,

$$\ell(X) \geq H(X) - \log H(X) - \log\log H(X) - \ldots - 6.$$

In this note we improve this bound to:

Theorem For every discrete random variable X,

$$\ell(X) \geq H(X) - \log(H(X)+1) - \log e. \qquad \square$$

This bound is asymptotically achieved by a random variable derived from the geometric distribution.

The next section proves these statements. The appendix recounts known proofs of Wyner's upper bound and of a lower bound that is generally weaker than the theorem's.

2 Proof

Without loss of generality, assume that $\mathcal{X} \subseteq \mathcal{N}$ $(= \{1, 2, \ldots\})$ and let $p_i \stackrel{\text{def}}{=} \Pr(X = i)$. Central to our proof is a (known) relation between $H(X)$ and

$$E(X) \stackrel{\text{def}}{=} \sum_{i \in \mathcal{X}} i p_i,$$

the expected value of X.

[2] The lower bound was later shown to hold for the larger class of *uniquely-decodable codes.*

Lemma 1 If X is distributed geometrically over $\mathcal{N}$ then

$$\log(E(X)) \leq H(X) \leq \log(E(X)) + \log e.$$

Proof: Suppose that X is distributed with parameter p: for $i \geq 1$, $p_i = p(1-p)^{i-1}$. Then

$$E(X) = \frac{1}{p},$$

while

$$H(X) = \log\frac{1}{p} + \frac{1-p}{p}\log\frac{1}{1-p} \leq \log\frac{1}{p} + \log e.$$

Note that

$$\frac{1-p}{p}\log\frac{1}{1-p} \geq (1-p)\log e,$$

hence the bound is asymptotically achievable as p decreases to 0. $\square$

Lemma 2 For every random variable X distributed over $\mathcal{N}$,

$$H(X) \leq \log(E(X)) + \log e.$$

Proof: Of all random variables distributed over $\mathcal{N}$ and having a given expectation, the entropy is maximized by a geometrically-distributed one, e.g., Cover and Thomas [5]. $\square$

A reverse type of the above inequality cannot hold. For every integer i, the constant random variable $X = i$ has zero entropy and expected value i. Even if the p_i's are required to be non-increasing, we can, for $E \geq 1$ and $m \geq 2(E-1)$, let $p_1 = 1 - \frac{2(E-1)}{m}$, and $p_2 = \ldots = p_m = \frac{2(E-1)}{m(m-1)}$. The resulting random variable has expectation E while its entropy diminishes to 0 with increasing m.

Lemma 3 For every discrete random variable X,

$$H(X) \leq \ell(X) + \log(\ell(X) + 1) + \log e.$$

Proof: xxx Without loss of generality, assume that X is distributed over $\mathcal{N}$, and that the sequence $p_i \stackrel{\text{def}}{=} \Pr(X = i)$ is non-increasing. Any encoding ϕ of X that achieves $\ell(X)$ has $|\phi(1)| = 0$, $|\phi(2)| = |\phi(3)| = 1$, and, in general,

$$|\phi(i)| = \lfloor \log i \rfloor.$$

xxx Motivated by the observation, we define the random variable

$$\Phi \stackrel{\text{def}}{=} |\phi(X)| = \lfloor \log X \rfloor.$$

By definition,

$$\ell(X) = \sum_{x\in\mathcal{X}} p(x)\phi(x) = E(\Phi).$$

To derive the theorem observe that Φ is a function of X, and that $|\phi^{-1}(\varphi)| = 2^{\varphi}$ for all $\varphi \in \{0, 1, \ldots\}$. Hence:

$$\begin{aligned}
H(X) &= H(\Phi) + H(X|\Phi) \\
&= H(\Phi) + \sum_{\varphi=0}^{\infty} p(\varphi) H(X|\Phi = \varphi) \\
&\leq H(\Phi) + \sum_{\varphi=0}^{\infty} p(\varphi) \log|\phi^{-1}(\varphi)| \\
&= H(\Phi) + \sum_{\varphi=0}^{\infty} p(\varphi)\cdot\varphi \\
&= H(\Phi) + E(\Phi) \\
&\leq E(\Phi) + \log\left(E(\Phi)+1\right) + \log e,
\end{aligned}$$

where the last inequality follows from Lemma 2 (slightly modified because Φ 'ranges' over $\{0, 1, \ldots\}$, rather than $\mathcal{N}$). □

Rephrased, this result gives a slightly stronger form of the theorem.

To show that the theorem's bound can be arbitrarily approximated, we 'reverse engineer' the proof of the last lemma. Take any $0 < p < 1$ and define the random varialbe $X = (Y, Z)$ where:

$$P(y) \stackrel{\text{def}}{=} p(1-p)^y \quad \text{for} \quad y \geq 0$$

and

$$P(z|y) = \begin{cases} \frac{1}{2^y} & \text{for } 2^j \leq i \leq 2^{j+1}-1 \\ 0 & \text{otherwise.} \end{cases}$$

That is,

$$P = \left(p, \frac{p(1-p)}{2}, \frac{p(1-p)}{2}, \frac{p(1-p)^2}{4}, \ldots\right).$$

Then if ϕ is an optimal encoding of X then $|\phi(x)| = y$, hence

$$\ell(X) = \sum_{j=0}^{\infty} j q_j = E(Y).$$

$$H(X) = H(Y) + H(Z|Y) = H(Y) + \sum_{y=0}^{\infty} p(y)\cdot y = H(Y) + E(Y) \geq E(Y) + \log(E(Y)+1) + (1-p)\log e$$

where the inequality follows from the remark ending Lemma 1's proof. Hence, as p decreases, $\ell(P)$ approaches $H(P) - \log(H(P)+1) - \log e$.

Appendix

For completeness, we recount known proofs of Wyner's upper bound and of a lower bound proven by Leung-Yan-Cheong and Cover [4].

The lower bound is generally weaker than the one claimed by the theorem, but its simplified proof, due to Dunham [6], is short and elegant:

Lemma 4 If $\mathcal{X}$ is finite, then

$$\ell(X) \geq H(X) - \log\log(|\mathcal{X}|+1).$$

Proof: An optimal code satisfies:

$$\sum_{i=1}^{|\mathcal{X}|} p_i \log\frac{1}{p_i} - \sum_{i=1}^{|\mathcal{X}|} p_i l_i = \sum_{i=1}^{|\mathcal{X}|} p_i \log\frac{2^{-l_i}}{p_i} \leq \log\sum_{i=1}^{|\mathcal{X}|} 2^{-l_i} = \log\sum_{i=1}^{|\mathcal{X}|} 2^{-\lfloor\log i\rfloor} \leq \log\log(|\mathcal{X}|+1). \quad \square$$

We note that Rissanen [7] showed that the $\log\log(|\mathcal{X}|+1)$ term in this bound can be slightly decreased.

To prove Wyner's upper bound, assume again that the p_i's are non-increasing. Then

$$p_i \leq \frac{1}{i}\ .$$

Taking an encoding ϕ where

$$|\phi(i)| = \lfloor\log i\rfloor \leq \log\frac{1}{p_i}\ ,$$

we obtain:

$$\ell(X) = \sum_{i\in\mathcal{X}} p_i|\phi(i)| \leq \sum_{i\in\mathcal{X}} p_i \log\frac{1}{p_i} = H(X).$$

This bound is trivially achieved by the constant random variables. For random variables with arbitrarily high entropy, it can be approached up to an additive constant of 2. Take $m = 2^n - 1$ and let X be uniformly distributed over $1,\ldots,m$. Then

$$H(X) = \log m$$

and

$$\ell(X) = \frac{1}{m}\sum_{i=0}^{n-1} i2^i = \frac{1}{m}((n-2)2^n+2) = \frac{1}{m}(n2^n - 2m) = \frac{n2^n}{m} - 2 \geq \log m - 2.$$

Acknowledgement

We thank Meir Feder for references to [2] and [7].

References

[1] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[2] A.D. Wyner. An upper bound on the entropy series. *Information and Control*, 20:176–181, 1972.

[3] P. Elias. Universal codeword sets and representations of the integers. *IEEE Transactions on Information Theory*, 21(2):194–203, March 1975.

[4] S.K. Leung-Yan-Cheong and T.M. Cover. Some equivalences between shannon entropy and kolmogorov complexity. *IEEE Transactions on Information Theory*, 24(3):331–338, May 1978.

[5] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[6] J.G. Dunham. Optimal noiseless coding of random variables. *IEEE Transactions on Information Theory*, 26(3):345, May 1980.

[7] J. Rissanen. Tight lower bounds for optimal code length. *IEEE Transactions on Information Theory*, 28(2):348–349, March 1982.