SOME INEQUALITIES BETWEEN SHANNON ENTROPY AND KOLMOGOROV, CHAITIN, AND EXTENSION COMPLEXITIES

BY

S.K. LEUNG-YAN-CHEONG and THOMAS M. COVER

TECHNICAL REPORT NO. 16
OCTOBER 10, 1975

PREPARED UNDER THE AUSPICES
OF
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
CONTRACT AFOSR #F44620-74-C-0068

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SOME INEQUALITIES BETWEEN SHANNON ENTROPY AND KOLMOGOROV, CHAITIN, AND EXTENSION COMPLEXITIES

by

S.K. LEUNG-YAN-CHEONG and THOMAS M. COVER

TECHNICAL REPORT NO. 16
OCTOBER 10, 1975

PREPARED UNDER THE AUSPICES
OF
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
CONTRACT AFOSR #F44620-74-C-0068

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Some Inequalities Between Shannon Entropy and Kolmogorov, Chaitin, and Extension Complexities

S.K. Leung-Yan-Cheong* and Thomas M. Cover**

Abstract

It is known that the expected code word length $L_{U,D}$ of the best uniquely decodable code satisfies $H(X) \leq L_{U,D} < H(X) + 1$. Let $X$ be a random variable which can take on $n$ values. Then it is shown that the average codeword length $L_{1:1}$ for the best 1:1 (not necessarily uniquely decodable) code for $X$ is shorter than the average codeword length $L_{U,D}$ for the best uniquely decodable code by no more than $(\log_2 \log_2 n) + 3$. Let $Y$ be a random variable taking on a finite or countable number of values and suppose it has entropy $H$. Then it is proved that $L_{1:1} \geq H - \log_2 (H+1) - \log_2 \log_2 (H+1) - \cdots - 4$. Some relations are established among the Kolmogorov, Chaitin and Extension complexities. Finally, it is shown that for all computable probability distributions, the universal prefix codes associated with the Chaitin complexity have expected codeword length within a constant of the Shannon entropy.

*Stanford University. The work of this author was supported under NSF Grants GK-33250 and ENG-10173.

**Stanford University. The work of this author was supported under AFOSR Contract F49620-74-C-0068.
I. Introduction.

Shannon has shown that the minimal expected length $L$ of a prefix code for a random variable $X$ satisfies

$$H(X) \leq L < H(X) + 1$$  \hspace{1cm} (1)

where $H$ is the entropy of the random variable. Shannon's restriction of the encoding or description of $X$ to prefix codes is highly motivated by the implicit assumption that the descriptions will be concatenated and thus must be uniquely decodable. Since the set of allowed code word lengths is the same for the uniquely decodable and instantaneous codes [1,2], the expected code word length $L$ is the same for both sets of codes. Shannon's result follows by assigning code word length $L_i = [\log_2 1/p_i]$ to the $i$-th outcome of the random variable, where $p_i$ is the probability of the $i$-th outcome. Thus the entropy $H$ plays a fundamental role and may be interpreted as the minimal expected length of the description of $X$. The intuition behind the entropy $H$ is so compelling that it would be disconcerting if $H$ did not figure prominently in a description of the most efficient coding with respect to other less constrained coding schemes. In particular we have in mind 1:1 codes, i.e., codes which assign a distinct binary code word to each outcome of the random variable, without regard to the constraint that concatenations of these descriptions be uniquely decodable. It will be shown here that $H$ is a first order approximation to the minimal expected length of 1:1 codes.
Throughout this paper we will use \( L_{1;1} \) and \( L_{U,D} \) to denote the average code word lengths for the best \( 1:1 \) code and uniquely decodable code respectively. Since the class of \( 1:1 \) codes contains the class of uniquely decodable codes, it is clear that in general \( L_{1;1} \leq L_{U,D} \). We shall show that \( L_{1;1} \geq H - \log \log n + c \) where \( n \) is the number of values that the random variable \( X \) can take on. Perhaps more to the point, we shall also show that \( L_{1;1} \geq H - \log(H+1) - O(\log \log(H+1)) \). Thus to first order a \( 1:1 \) code allows no more compression than a uniquely decodable or prefix code.

As a consequence of the work of Kolmogorov and Chaitin, a notion of the intrinsic descriptive complexity of a finite object has been developed. This is closely related to the work of Shannon in which the complexity of a class of objects is defined in terms of the probability distribution over that class. The complexity measures of Kolmogorov and Chaitin, together with a new complexity measure which we call the extension complexity, have associated with them universal coding schemes. We shall establish that a universal encoding has an expected codeword length with respect to any probability distribution on the set of possible outcomes which is within a constant of the Shannon entropy, thus establishing a tie between the individual complexity measure of Kolmogorov and the average complexity measure of Shannon.

In section II, we consider a random variable which can take on only a finite number of values and we maximize \( (L_{U,D} - L_{1;1}) \). In section III we derive lower bounds on \( L_{1;1} \) in terms of the entropy of a random variable taking values in a countable set. In section IV we recall the
definitions of the Kolmogorov and Chaitin complexities of binary sequences and introduce the notion of an Extension complexity. We then derive some relationships between these quantities. Finally, in section V, we show that for all computable probability distributions the universal prefix codes associated with the Chaitin complexity have expected codeword length within a constant of the Shannon entropy.

II. Maximization of \( (I_{U,D} - L_{1:1}) \).

Let \( X \) be a random variable taking on a finite number of values, i.e.,

\[
X = \begin{pmatrix}
X_1 & X_2 & \cdots & X_n \\
P_1 & P_2 & \cdots & P_n
\end{pmatrix}.
\]

With no loss of generality assume \( p_1 \geq p_2 \geq \cdots \geq p_n \). Let \( \ell_i \), \( i = 1, 2, \ldots, n \) be the lengths of the codewords in the best 1:1 code for encoding the r.v. \( X \), where \( \ell_i \) is the length of the codeword assigned to \( X_i \).

Remark: Unless otherwise stated, all logarithms are to the base 2.

It is clear that the best 1:1 code must have \( \ell_1 \leq \ell_2 \leq \ell_3 \leq \cdots \).

Thus, by inspection,

\[
\ell_i = \lceil \log(\frac{1}{2} + 1) \rceil
\]

and

\[
L_{1:1} = \sum_{i=1}^{n} p_i \ell_i = \sum_{i=1}^{n} p_i \left\lceil \log(\frac{1}{2} + 1) \right\rceil.
\]
We now prove the following theorem which gives an upper bound on
\( (L_{U,D} - L_{1:1}) \).

**Theorem 1:**

\[
L_{1:1} \geq L_{U,D} - \log \log n - 3.
\]  

(4)

**Proof:**

From (1) we have \( L_{U,D} < H(X) + 1 \). Therefore

\[
\max(L_{U,D} - L_{1:1}) < 1 + \max(H(X) - L_{1:1}).
\]  

(5)

Noting from (3) that

\[
L_{1:1} \geq \sum_{i=1}^{n} p_i \log \left( \frac{1}{2} + 1 \right)
\]  

(6)

we can write

\[
H(X) - L_{1:1} \leq \sum_{i=1}^{n} p_i \left( \log \frac{1}{p_i} - \log \left( \frac{1}{2} + 1 \right) \right).
\]  

(7)

We then use the method of Lagrange multipliers to maximize the right hand side of (7). The proof is completed by using (5). Details of the proof are given in Appendix A.

III. Lower bounds on \( L_{1:1} \) in terms of the entropy \( H \).

The objective in this section is to obtain lower bounds on \( L_{1:1} \) in terms of the entropy \( H \) of the random variable. As a first step we consider transformations of \( 1:1 \) to U.D. codes. The random variables
considered may take on a countable number of values.

Some possible transformations from 1:1 to U.D. codes.

The aim here is to find efficient means of transforming 1:1 codes
to U.D. codes.

Let \( \ell_1, \ell_2, \ldots \) be the lengths of the codewords for the best 1:1
code; assume \( \ell_1 \leq \ell_2 \leq \cdots \).

Let \( f \) be any function such that \( \sum_{i}^{-f(\ell_i)} \leq 1 \). Then from Kraft's
inequality, the set of lengths \( \{ f(\ell_1) \} \) yields acceptable word lengths
for a prefix (or U.D.) code. If \( f \) is integer-valued and \( \sum_{i}^{-f(\ell_i)} > 1 \),
\( \{ f(\ell_i) \} \) cannot yield a prefix code.

Theorem 2:

The following functions represent possible transformations from 1:1
to U.D. codes.

(i) \( f(\ell_i) = \ell_i + a \left\lceil \log \ell_i \right\rceil + c \) where \( a > 1 \) and \( c \geq \log \left( \frac{a}{a-1} \right) \) (8)

(ii) \( f(\ell_i) = \ell_i + 2 \left\lceil \log (\ell_i + 1) \right\rceil \) \hspace{1cm} (9)

(iii) \( f(\ell_i) = \ell_i + \left\lceil \log \ell_i \right\rceil + \left\lceil \log (\log \ell_i) \right\rceil + \cdots + 2 \) \hspace{1cm} (10)

The proof of Theorem 2 follows from verification of the Kraft inequality for
\( f(\ell_i) \) and is given in Appendix B.

We now make use of Theorem 2 to prove some lower bounds on \( L_{1:1} \) in
terms of the entropy \( H \).
Theorem 3:

The expected length $L_{1:1}$ of the best $1:1$ code satisfies the following lower bounds:

i) $L_{1:1} \geq H - a(1 + \log(H+1)) - \log\left(\frac{a}{a-1}\right)$ where $a > 1$, $c \geq \log\left(\frac{a}{a-1}\right)$ (11)

ii) $L_{1:1} \geq H - 2 \log(H+2)$ (12)

iii) $L_{1:1} \geq H - \log(H+1) - \log \log(H+1) - \ldots - \frac{3}{4}$ (13)

Proof:

(i) From Theorem 2(i) and the fact that the expected length for a U.D. code $\geq H(X)$ we can write

$$E(\ell + a \, \left\lceil \log \ell \right\rceil + c) \geq H \text{ where } a > 1, \ c \geq \log\left(\frac{a}{a-1}\right).$$

Therefore $E\ell + a(1 + E \log \ell) + c \geq H$ where $E\ell = L_{1:1}$. From Jensen's inequality and the convexity of $\ell - \log \ell$ we have $E\ell + a + a \log E\ell + c \geq H$. But $E\ell < H + 1$, since $\ell$ corresponds to the best $1:1$ code, which is certainly better than the best prefix code, and we know that the expected length for the best prefix code is less than $(H+1)$. Thus

$$E\ell \geq H - a(1 + \log(H+1)) - \log\left(\frac{a}{a-1}\right)$$

(ii) From Theorem 2(ii) and the fact that $L_{U.D.} \geq H$, we have

$$E(\ell + 2 \left\lceil \log(\ell+1) \right\rceil) \geq H$$

$$E\ell + 2 E \log(\ell+1) \geq H$$
By Jensen's inequality, \( \mathbb{E} \ell + 2 \log(\mathbb{E} \ell + 1) \geq H \).

But \( \mathbb{E} \ell < H+1 \) as before. Thus

\[
\mathbb{E} \ell + 2 \log(H+2) \geq H
\]

\[
L_{\ell+1} \geq H - 2 \log(H+2) .
\]

(iii) From Theorem 2(iii) and the fact that \( L_{\text{U.D.}} \geq H \), we have

\[
\mathbb{E} (\ell + \log \ell + \log(\log \ell) + \cdots + 2) \geq H . \tag{14}
\]

Thus

\[
\mathbb{E} (\ell + \log \ell + \log(\log \ell) + \cdots + 2) \geq H . \tag{15}
\]

For convenience we will define the function \( \log^* n \) by \( \log^* n \triangleq \log n + \log \log n + \cdots \) stopping at the last positive term.

Then

\[
\mathbb{E} (\ell + \log^* \ell + 2) \geq H . \tag{16}
\]

However, \( \log^* \ell \) is not concave. But in Appendix C, we prove that there exists a (piecewise-linear) concave function \( F^*(\ell) \) such that

\[
F^*(\ell) \leq \log^* \ell < F^*(\ell) + 2 . \tag{17}
\]

This, in conjunction with (15), enables us to obtain

\[
\mathbb{E} \ell + \log^* \mathbb{E} \ell + \frac{1}{2} \geq H . \tag{18}
\]
But $E \varepsilon < H+1$ as before. Therefore

$$L_{1,1} \geq H - \log(H+1) - \log \log(H+1) - \cdots - 4.$$  (19)

IV. Some relations between Kolmogorov, Chaitin and Extension complexities.

Let $(0,1)^*$ denote the set of all binary finite length sequences, including the empty sequence. For any $x = (x_1, x_2, \ldots) \in (0,1)^* \cup (0,1)^\infty$, let $x(n) = (x_1, x_2, \ldots, x_n)$ denote the first $n$ bits of $x$.

**Definition:** A subset $S$ of $(0,1)^*$ is said to have the prefix property if and only if no sequence in $S$ is the proper prefix of any other sequence in $S$.

Thus, for example, {00,100} has the prefix property, but {00,001} does not.

**Definition:** The Kolmogorov complexity of a binary sequence $x(n) \in (0,1)^n$ with respect to a partial recursive function $A: (0,1)^* \times N \to (0,1)^*$ is defined to be

$$K_A(x(n)|n) = \min_{A(p,n)=x(n)} \ell(p)$$  (20)

where $\ell(\cdot)$ is the length of the sequence $p$.

Here $A$ may be considered to be a computer, $p$ its program, and $x$ its output. We shall use interchangeably the recursive function theoretic terminology and computer terminology. (See, for example, Chaitin [3] for a discussion of the equivalence of the two).
Definition: Let $U: \{0,1\}^* \to \{0,1\}^*$ be a partial recursive function with a prefix domain. Then the Chaitin complexity of a binary sequence $x$ with respect to $U$ is given by

$$C_U(x) = \min_{U(p)=x} \ell(p).$$

(21)

We now introduce a new complexity measure.

Definition: Let $U: \{0,1\}^* \to \{0,1\}^*$ be a partial recursive function with a prefix domain. Then the Extension complexity of a binary sequence $x$ with respect to $U$ is defined by

$$E_U(x) = \min_{U(p) \supseteq x} \ell(p)$$

(22)

where $U(p) \supseteq x$ means that $U(p)$ is an extension of $x$.

Definition: Given a complexity measure $C_B^*: \Omega \to \mathbb{N}$ where $\Omega$ is countable and $B$ is a partial recursive function, we shall say that $C^*(\cdot)$ is universal if there exists a partial recursive function $U_0$ such that for any other partial recursive function $A$,

$$\exists \text{ constant } c, \forall \omega \in \Omega, C^*_{U_0}(\omega) \leq C^*_A(\omega) + c$$

(23)

It has been shown [3,4] that the Kolmogorov and Chaitin complexity measures are universal. The same result can be shown to hold for the Extension complexity measure. Thus from now on, we will assume that the complexities are measured with respect to some fixed appropriate universal function and subscripts will be dropped. We shall denote the Chaitin, Kolmogorov and Extension complexities of a binary sequence $x \in \{0,1\}^*$ by $C(x)$, $K(x|\ell(x))$ and $E(x)$ respectively.
Theorem 4:

There exist constants $c_0$ and $c_1$ such that $\forall x \in \{0,1\}^*$,

$$E(x) + c_0 \leq C(x) \leq E(x) + \log \ell(x) + \log \log \ell(x) + \cdots + c_1 \quad (25)$$

Proof:

The first inequality follows directly from the definitions of $E(x)$ and $C(x)$.

To prove the second inequality, note that the Chaitin complexity program $p'$ can be constructed from the Extension complexity program $p$ as follows.

Let $s$ be the shortest program (from a set having the prefix property) for calculating $\ell(x)$. Then $p'$ is the concatenation $qsp$ where $q$ consists of a few bits to tell the computer to expect 2 programs and interpret them appropriately. So we have

$$C(x) \leq E(x) + C(\ell(x)) + c_2 \quad (26)$$

From Theorem 2(iii)

$$C(\ell(x)) \leq \log \ell(x) + \log \log \ell(x) + \cdots + c_3 \quad (27)$$

Combining (26) and (27) yields Theorem 4.

Theorem 5:

There exist constants $c_0$ and $c_1$ such that $\forall x \in \{0,1\}^*$,

$$K(x|\ell(x)) + c_0 \leq C(x) \leq K(x|\ell(x)) + \log K(x|\ell(x)) + \cdots + \log \ell(x) + \log \log \ell(x) + \cdots + c_1 \quad (28)$$

Proof:

The first inequality is a direct consequence of the definitions.
To prove the second inequality, we first note that the Chaitin complexity measure is defined with respect to a computer whose programs belong to a set with the prefix property. From Theorem 2(iii), we know that we can transform the domain of a Kolmogorov complexity measure computer into one which has the prefix property by extending the length of the Kolmogorov complexity program from \( K(x|\ell(x)) \) to 
\[
K(x|\ell(x)) + \log_2 K(x|\ell(x)) + \cdots + c_2.
\]
Let us denote this extended program by \( \mathbf{p} \). From the proof of Theorem 4, we also know that a program \( s \) (belonging to a set with the prefix property) which describes the length of \( x \) need not be longer than 
\[
\log_2 \ell(x) + \log_2 \log_2 \ell(x) + \cdots + c_2.
\]
The Chaitin complexity program can be the concatenation \( q \ast p \) where \( q \) consists of a few bits to tell the computer to expect 2 programs and interpret them appropriately. So
\[
C(x) \leq K(x|\ell(x)) + \log_2 K(x|\ell(x)) + \cdots \\
+ \log_2 \ell(x) + \log_2 \log_2 \ell(x) + \cdots + c_2.
\]

This completes the proof of Theorem 5.

V. Relation of Chaitin code length to Shannon code length.

Let \( \{X_i\}_{i=1}^\infty \) be a stationary binary stochastic process with marginals \( p(x(n)), x(n) \in \{0,1\}^\infty, n = 1,2,\ldots \) and Shannon entropy
\[
H(X) = \lim_{n \to \infty} \frac{H(X_1, X_2, \ldots, X_n)}{n}. \tag{29}
\]

Let
\[
C(x(n)|n) = \min_{U(p,n^*) = x(n)} \ell(p). \tag{30}
\]
be the Chaitin complexity of $x(n)$ given $n^*$, the shortest length binary program for $n$ (See Chaitin [3] for definitions of conditional complexities). As before, the domain of $U(\cdot, n^*)$ has the prefix property, for each $n$.

The Shannon entropy $H(x_1, \ldots, x_n)$ is a real number, while the Chaitin complexity $C(x_1, \ldots, x_n | n)$ is a random variable equal to the length of the shortest codeword (program) assigned to $(x_1, \ldots, x_n)$ by $U$. The prefix set of codewords so defined may be thought of as a universal prefix encoding of $n$-sequences, for each $n$. Note in particular that the prefix encoding induced by $U$ is completely oblivious to the true underlying statistics $p(x_1, \ldots, x_n)$. We shall show, however, that this universal encoding has an expected word length equal to first order to the optimal Shannon bound $H(x_1, \ldots, x_n)$.

**Theorem 6:**

For every computable probability measure $p: \{0, 1\}^* \rightarrow [0, 1]$ for a stochastic process, there exists a constant $c$, such that for all $n$

$$H(x_1, \ldots, x_n) \leq E_p C(x_1, \ldots, x_n | n) \leq H(x_1, \ldots, x_n) + c$$

(31)

**Proof:**

Clearly, for each $n$, $C(x(n) | n)$, $x(n) \in \{0, 1\}^n$ must satisfy the Kraft inequality. So we have

$$H(x_1, \ldots, x_n) \leq E_p C(x_1, \ldots, x_n | n).$$

(32)

For the right half of the inequality, we must use a theorem of Chaitin relating $C$ and a certain universal probability measure $P^*$. 

12
We then relate $P^*$ to the true statistics $P$ to achieve the desired proof. We define

$$P^*(x(n) | n) = \sum_{U(p, n^*)=x(n)} 2^{-\ell(p)}. \tag{33}$$

Chaitin has shown [3, Theorem 3.5] that there exists a constant $c'$ such that

$$C(x(n) | n) \leq \log \frac{1}{P^*(x(n) | n)} + c' \tag{34}$$

for all $n$. In addition, he has shown that for any other prefix domain computer $A$, there exists a constant $c''$ such that

$$P^*(x(n) | n) \geq c'' P_A(x(n) | n) \tag{35}$$

for all $n$, where $P_A(\cdot)$ is defined as in (33).

Below, in Lemma 1, we show that for the given computable statistics $p: \{0,1\}^* \to [0,1]$ that there exists a prefix domain computer $A$ such that $P_A(x(n) | n) = p(x(n))$, for all $n$. The proof can then be completed as follows:

$$E C(x(n) | n) = \sum_p \sum_{x(n) \in \{0,1\}^n} p(x(n)) C(x(n) | n) \tag{36}$$
\[ \leq \sum_{x(n) \in \{0,1\}^n} p(x(n)) \left( \log \frac{1}{P_A(x(n)|n)} + c' \right) \text{, using (34)}, \quad (37) \]

\[ \leq \sum_{x(n) \in \{0,1\}^n} p(x(n)) \left( \log \frac{1}{c'' P_A(x(n)|n)} \right) + c', \text{ using (35)}, \quad (38) \]

\[ = \sum_{x(n) \in \{0,1\}^n} p(x(n)) \log \frac{1}{p(x(n))} + c''', \text{ using Lemma 1}, \quad (39) \]

\[ = H(X_1, \ldots, X_n) + c''', \quad \forall \ n. \quad (40) \]

Q.E.D.

**Lemma 1:**

For given computable statistics \( p : \{0,1\}^* \rightarrow [0,1] \) there exists a prefix domain computer \( A \) such that \( P_A(x(n)|n) = p(x(n)), \ \forall \ n. \)

**Proof:**

Since \( p(x(n)) \) is computable, this implies

\[ F(x(n)) \triangleq \sum_{x'(n) < x(n)} p(x'(n)) \quad (41) \]

is computable, where \( x'(n) < x(n) \) means \( x'(n) \) precedes \( x(n) \) in a lexicographic ordering of the \( n \)-sequences.

Let \( A \) be a computer that has \( n^* \) on its work tape. \( A \) first calculates \( n \), then inspects a random program \( p \in \{0,1\}^{\infty} \) until it is
sure that \( p \in (\cdot F(x(n)), \cdot F(x(n)+00\ldots 001)) \) for some \( x(n) \in \{0,1\}^n \),

where \( p = \sum_{i=1}^{\infty} (1/2)^i p_i \), and \( x(n)+00\ldots 001 \) means the sequence obtained by adding \( x(n) \) and \( (1/2)^n \) and reinterpreting as a sequence. Finally, A prints out this \( x(n) \).

It is easily seen that

\[
\Pr\{p \in (\cdot F(x(n)), \cdot F(x(n)+00\ldots 001))\} = p(x(n)),
\]

\( \forall x(n) \in \{0,1\}^n \). This completes the construction.

VI. Conclusions.

This study can be perceived in three parts. First, the minimal average code length with respect to known statistics has been shown to be equal to the Shannon entropy \( H \) to first order under different coding constraints. Second, the individual complexity measures of Kolmogorov, Chaitin, and others have been shown to be equivalent to one another, also to first order. Finally, the expected code length of the individual complexity measure has been shown to be equal to first order to the Shannon entropy, thus tying these approaches together.

Acknowledgement:

The authors would like to thank Professor John T. Gill for suggesting the method used for lower bounding \( l_{1:1} \) in section III.
Appendix A. Proof of Theorem 1.

Theorem 1:

\[ L_{1.1} \geq L_{U.D.} - \log \log n - 3. \]

Proof:

From (1)

\[ \max (L_{U.D.} - L_{1.1}) < 1 + \max (H(X) - L_{1.1}). \] \hspace{1cm} (A1)

We now proceed to find \( \max (H(X) - L_{1.1}) \). Let \( A \triangleq H(X) - L_{1.1} \).

Then

\[ A = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} - \sum_{i=1}^{n} p_i \left[ \log \left( \frac{i}{2} + 1 \right) \right] \] \hspace{1cm} (A2)

\[ \leq \sum_{i=1}^{n} p_i \left[ \log \frac{1}{p_i} - \log \left( \frac{i}{2} + 1 \right) \right] \] \hspace{1cm} (A3)

\[ \max A \leq \max \sum_{i=1}^{n} p_i \left[ \log \frac{1}{p_i} - \log \left( \frac{i}{2} + 1 \right) \right]. \] \hspace{1cm} (A4)

Let

\[ c_i \triangleq \ln \left( \frac{i}{2} + 1 \right) \] \hspace{1cm} (A5)

Let

\[ J(p_1, \ldots, p_n) \triangleq \sum_{i=1}^{n} p_i \left[ \ln \frac{1}{p_i} - c_i \right] + \lambda \sum_{i=1}^{n} p_i. \] \hspace{1cm} (A6)

Differentiating \( J(p_1, \ldots, p_n) \) with respect to \( p_i \), we obtain
\[ \frac{\partial J}{\partial p_i} = -c_i + \lambda - 1 + \ln \frac{1}{p_i} . \quad (A7) \]

Setting \( \frac{\partial J}{\partial p_i} = 0 \), we obtain

\[ \ln p_i = \lambda - (c_i + 1) \quad (A8) \]

i.e. \( p_i = e^{\lambda - (c_i + 1)} = a e^{-c_i} \quad (A9) \)

where \( a \) is some constant.

Now

\[ \sum_{i=1}^{n} p_i = 1 . \quad (A10) \]

Substituting (A9) in (A10) and using (A5) we get

\[ 2a \sum_{i=1}^{n} \frac{1}{i+2} = 1 . \quad (A11) \]

Let

\[ H_k \triangleq \sum_{i=1}^{k} \frac{1}{i} . \quad (A12) \]

Then (A11) can be rewritten as

\[ 2a \left( H_{n+2} - H_2 \right) = 1 . \quad (A13) \]

From (A9) and (A5)

\[ p_i = \frac{2a}{i+2} . \quad (A14) \]
Therefore,

\[
\max \sum_{i=1}^{n} p_i \left( \ln \frac{1}{p_i} - \ln \left( \frac{1}{2} + 1 \right) \right)
\]

\[
= \sum_{i=1}^{n} \frac{2a}{i+2} \left( \ln \left( \frac{i+2}{2a} \right) - \ln \left( \frac{i+2}{2} \right) \right)
\]

\[
= \ln \left( \frac{1}{a} \right) \cdot 2a \sum_{i=1}^{n} \frac{1}{i+2}
\]

\[
= \ln \left( \frac{1}{a} \right) . \quad (A15)
\]

So from (A14)

\[
\max A \leq \log \left( \frac{1}{a} \right) = 1 + \log \left( H_{n+2} - H_2 \right) . \quad (A16)
\]

Using (A1) we obtain

\[
\max (L_{U.D.} - L_{1:1}) < 2 + \log \left( H_{n+2} - H_2 \right) \quad (A17)
\]

Knuth [5] has

\[
H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \epsilon , \quad (A18)
\]

where \(0 < \epsilon < \frac{1}{252n}\) and \(\gamma = 0.577...\) is Euler's constant. Therefore,
\[ \max(L_{U.D.} - L_{1:1}) < 2 + \log(\ln(n+2)) \] (A19)

\[ \leq 3 + \log \log n . \] (A20)

Thus we conclude that

\[ L_{1:1} \geq L_{U.D.} - \log \log n - 3 . \] (A21)
Appendix B. Admissible lengths for uniquely decodable codes.

In this appendix, we prove Theorem 2 which states that the following functions represent possible transformations from 1:1 to U.D. codes. Recall $l_1 = \lceil \log(\frac{a}{a-1}) \rceil$.

(i) $f(l_1) = l_1 + a \lceil \log l_1 \rceil + c$ where $a > 1$ and $c \geq \log(\frac{a}{a-1})$ (Bl)

(ii) $f(l_1) = l_1 + 2 \lceil \log(l_1+1) \rceil$ (B2)

(iii) $f(l_1) = l_1 + \lceil \log l_1 \rceil + \lceil \log(\log l_1) \rceil + \cdots + 2$ (B3)

Proof of (i). Define

$$S = \sum_{i=1}^{\infty} 2^{-f(l_1)} = \sum_{i=1}^{\infty} 2^{-f(l_1)} = \sum_{i=1}^{\infty} 2^{-l_1 - a \lceil \log l_1 \rceil - c} .$$

But there are $2^k$ 1:1 codewords of length $k$. Therefore

$$S = 2^{-c} \sum_{\ell=1}^{\infty} \frac{1}{2^a \ell \lceil \log \ell \rceil}$$

$$= 2^{-c} \left( \frac{1}{2^0} + \frac{1}{2^a} \cdot 1 + \frac{1}{2^{2a}} \cdot 2^1 + \frac{1}{2^{3a}} \cdot 2^2 + \cdots + \frac{1}{2^{ka}} \cdot 2^{k-1} + \cdots \right)$$
If \( a = 1 \), \( S = 2^{-c} \left\{ 1 + \frac{1}{2} + \frac{1}{2} + \cdots \right\} = \infty \).

So \( f(\ell_1) = \ell_1 + \lfloor \log \ell_1 \rfloor + c \), where \( c \) is a fixed constant, cannot yield a prefix code.

For \( a > 1 \), \( S \) can be bounded for fixed \( c \).

Proof:

\[
S < 2^{-c} \sum_{\ell=1}^{\infty} \frac{1}{2^a \log \ell} = \sum_{\ell=1}^{\infty} \frac{1}{\ell^a}.
\]

But \( I < \sum_{\ell=1}^{\infty} \frac{1}{\ell^a} < I+1 \) where \( I = \frac{1}{a-1} \) because

\[
\int_1^{\infty} \frac{1}{x^a} \, dx = \left. \left( \frac{x^{1-a}}{1-a} \right) \right|_1^{\infty} = \left( \frac{1}{1-a} \right) = \frac{1}{a-1} = I.
\]

Therefore,

\[
S < 2^{-c} \left( \frac{1}{a-1} + 1 \right) = 2^{-c} \left( \frac{a-1}{a-1} \right).
\]

To make \( S \leq 1 \), it is sufficient to have \( c \geq \log \left( \frac{a}{a-1} \right) \). This completes the proof of (i).

Proof of (ii).

In this case, define
\[ S = \sum_{i=1}^{\infty} \frac{-f(\ell_i)}{2} \]

\[ = \sum_{i=1}^{\infty} \frac{-\ell_i}{2} - \frac{1}{2} \log(\ell_i + 1) \]

\[ = \sum_{k=1}^{\infty} \frac{1}{2^{\log(\ell+1)}} \], using the fact that there are \(2^k\) 1:1 codewords of length \(k\),

\[ = \left( \frac{1}{2^2} \right) \cdot 2 + \left( \frac{1}{2^{2 \cdot 2}} \right) 2^2 + \left( \frac{1}{2^{2 \cdot 3}} \right) 2^3 \ldots + \left( \frac{1}{2^{2 \cdot k}} \right) 2^k + \ldots \]

\[ = \frac{1}{2^2} + \frac{1}{2^3} + \ldots + \frac{1}{2^k} + \ldots \]

\[ = \frac{1}{2^{\log(1 - \frac{1}{2})}} = 1. \]

This proves (ii).

Proof of (iii).

Let \( f(\ell_i) = \ell_i + \left\lfloor \log \ell_i \right\rfloor + \left\lfloor \log(\log \ell_i) \right\rfloor + \ldots + c \) where it is understood that we only consider the first \( k \) iterates for which \( \left\lfloor \log(\log(\ldots(\log \ell_i)\ldots)) \right\rfloor \) is positive,

\[ \text{e.g. if } \ell_i = 2, \ f(\ell_i) = 2 + 1 + c = 3 + c \]

\[ \ell_i = 5, \ f(\ell_i) = 5 + 2 + 1 + c = 8 + c. \]
Now

\[ S \triangleq \sum_{i=1}^{\infty} 2^{-f(\ell_i)} \]

\[ = 2^{-c} \sum_{i=1}^{\infty} \ell_i \cdot \left[ -\log \ell_i \right] \cdot \left[ -\log(\log \ell_i) \right] \ldots \]

\[ = 2^{-c} \left( \sum_{\ell=1}^{\infty} 2^{-\left[ -\log \ell \right]} \cdot 2^{-\left[ -\log(\log \ell) \right]} \ldots \right), \quad \text{since there} \]

\[ \quad \text{are} \ 2^k \ \text{1:1 codewords of length} \ k, \]

\[ = 2^{-c} \sum_{\ell=1}^{\infty} \frac{1}{2^{\left[ -\log \ell \right]}} \cdot \frac{1}{2^{\left[ -\log(\log \ell) \right]}} \ldots \]

\[ = 2^{-c} \left( \frac{1}{2^0} + \left( \frac{1}{2^1} \cdot \frac{1}{2^0} \right)2^1 + \left( \frac{1}{2^2} \cdot \frac{1}{2^1} \cdot \frac{1}{2^0} \right) \cdot 2^2 + \ldots \right. \]

\[ + \left( \frac{1}{2^k} \cdot \frac{1}{2^{k-1}} \ldots \frac{1}{2^0} \right)2^k + \ldots \right) \]

\[ = 2^{-c} \left( 1 + 1 + \frac{1}{2} + \frac{1}{2^2} \cdot 2 + \ldots + \frac{1}{2^{(k-1)k/2}} \ldots \right) \]

\[ < 2^{-c} \left( 1 + 1 + \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{4}} \right) \]

\[ < 3 \cdot 2^{-c}. \]

If we choose \( c = 2 \), then \( S \leq 1 \).

This completes the proof for (iii).

Remark: By exactly the same method used for proving (iii), it can easily be shown that

\[ f(\ell_i) = \ell_i + \left[ \log \ell_i \right] + \left[ \log(\log \ell_i) \right] + \ldots + 1 \]

is also a permissible transformation.
Appendix C.

In this Appendix, we exhibit a piecewise-linear concave function $F^*(\ell)$ such that

\[ F^*(\ell) \leq \log^* \ell < F^*(\ell)+2, \quad \ell \geq 1. \]  

(C1)

Recall $\log^* \ell = \log \ell + \log \log \ell + \ldots$ stopping at the last positive term. (See Figure 1)

The function $F^*(\ell)$ is also sketched in Figure 1. For $1 \leq \ell \leq 4$, $F^*(\ell) = \ell - 1$ and for $\ell > 4$, $F^*(\ell)$ is defined as follows: Consider the following sequence of values for $\ell$: $4, 4\sqrt{2}, 8, 8\sqrt{2}, 16, \ldots$, i.e., a geometric sequence with a ratio of $\sqrt{2}$. $F^*(\ell)$ is obtained by joining adjacent points on the $\log^* \ell$ curve by straight line segments for the $\ell$ values mentioned above.

In the following, for notational convenience we shall use

\[ \exp_2^r(x) = 2^{2^\ldots^2} \text{ r times} \quad \text{and} \quad \log_2^r(\ell) = \log \log \ldots \log \ell \quad \text{i.e. the} \quad \text{r-fold composition of the exponential and log functions, respectively.} \]

First, we prove the concavity of $F^*(\ell)$, $\ell > 1$.

Let us look at $F^*(\ell)$ for $\ell > 4$. It is clearly sufficient to prove concavity at points $\exp_2^r(2)$, $r = 3, 4, \ldots$ since concavity is automatically satisfied at all other points. Thus we need to show that
Figure 1. Sketch of $\log^* \ell$ and $F^*(\ell)$. 
\[
\frac{f(I) - f\left(\frac{I}{\sqrt{2}}\right)}{I - \frac{I}{\sqrt{2}}} \geq \frac{f(\sqrt{2} I) - f(I)}{\sqrt{2} I - I}, \quad I = \exp_{2}^{(r)}(2), \quad r = 3, 4, \ldots
\]  
(C2)

i.e. \[
\sqrt{2} \left| f(I) - f\left(\frac{I}{\sqrt{2}}\right) \right| \geq f(\sqrt{2} I) - f(I)
\]
(C3)

where for convenience we have set \( \log^{*} \ell = f(\ell) \). By definition,

\[
f(I) = \log I + \log \log I + \cdots + \log^{(r)}(I)
\]
(C4)

\[
f\left(\frac{I}{\sqrt{2}}\right) = \left(\log I - \frac{1}{2}\right) + \log \left(\log I - \frac{1}{2}\right) + \cdots + \log^{(r-1)}(\log I - \frac{1}{2})
\]
(C5)

\[
f(\sqrt{2} I) = \left(\log I + \frac{1}{2}\right) + \log \left(\log I + \frac{1}{2}\right) + \cdots + \log^{(r-1)}(\log I + \frac{1}{2})
\]

\[+ \log^{(r)}(\log I + \frac{1}{2}) . \]
(C6)

Consider the \(1^{\text{st}}\) terms in \( f(I) \), \( f\left(\frac{I}{\sqrt{2}}\right) \) and \( f(\sqrt{2} I) \)

\[
\left| f(I) - f\left(\frac{I}{\sqrt{2}}\right) \right|_{1^{\text{st}} \text{term}} = \frac{1}{2}
\]
(C7)

\[
\left| f(\sqrt{2} I) - f(I) \right|_{1^{\text{st}} \text{term}} = \frac{1}{2} . \]
(C8)

So considering the \(1^{\text{st}}\) terms only, (C5) is satisfied. In fact the difference between the L.H.S. and R.H.S. of (C5) is \( \frac{\sqrt{2} - 1}{2} = 0.207 \).
Now consider the 2nd terms in \( f(I), f(\frac{I}{\sqrt{2}}), f(\sqrt{I}) \). Because the log function is concave it is clear (see Figure 2) that

\[
\log \log I - \log \left( \log I - \frac{1}{2} \right) > \log \left( \log I + \frac{1}{2} \right) - \log \log I \tag{C9}
\]

Figure 2. Graphical interpretation of inequality (C9): \( \alpha > \beta \)

So considering the 2nd terms only of \( f(I), f(\frac{I}{\sqrt{2}}), f(\sqrt{I}) \) we see that (C3) is again satisfied. It is clear that by the same argument as above, the 3rd through rth terms of \( f(I) - f(\frac{I}{\sqrt{2}}) \) exceed the corresponding terms of \( f(\sqrt{I}) - f(I) \). There is one remaining term in \( f(\sqrt{2} I) \) which we
have to consider, namely $\log^{(r+1)}(\sqrt{2} I) \triangleq g(r)$. We now show that $g(r)$ is monotone decreasing in $r, r \geq 1$

$$g(r) = \log^{(r+1)}(\sqrt{2} \exp_2^r(2))$$

(C10)

$$= \log^{(r+2)}(2^{\sqrt{2} \exp_2^r(2)})$$

(C11)

$$g(r+1) = \log^{(r+2)}(\sqrt{2} \exp_2^{r+1}(2))$$

(C12)

So we need to show

$$(\exp_2^{r+1}(2))^{\sqrt{2}} \sqrt{2} \exp_2^{r+1}(2)$$

(C13)

i.e. $$(\exp_2^{r+1}(2))^{\sqrt{2}-1} > \sqrt{2}$$

(C14)

which is clearly satisfied for $r \geq 1$.

By inspection $\log \log \log \log 16 \sqrt{2} = 0.16$ which is less than $\frac{\sqrt{2}-1}{2}$ so that (C3) is satisfied for $I = \exp_2^r(2), r \geq 3$. To complete the proof of the convexity of $F^*(\ell)$, it can easily be verified that convexity of $F^*(\ell)$ also holds at $\ell = 4$.

We now proceed to show that

$$F^*(\ell) \leq f(\ell) < F^*(\ell) + 2, \text{ for } \ell \geq 1.$$

(C15)
Define an auxiliary function $a(\ell) \triangleq \frac{1}{4} \log \ell$. Consider the derivative $f'(\ell)$ of $f(\ell)$. If $\exp_2^{(r)}(2) \leq \ell < \exp_2^{(r+1)}(2)$, then

$$f'(\ell) = \frac{\log e}{\ell} + \frac{\log e}{\log \ell} \cdot \frac{\log e}{\ell} + \ldots + \frac{\log e}{\log^{(r)}(\ell)} \cdot \frac{\log e}{\log^{(r-1)}(\ell)} \ldots \frac{\log e}{\ell}$$

(C16)

We will now show that

$$f'(\ell) < a'(\ell) \text{ for } r \geq 2$$

(C17)

i.e. $f'(\ell) < \frac{1}{\ell} \cdot \log e$

(C18)

i.e. $\frac{\log e}{\log \ell} + \frac{\log e}{\log \log \ell} \cdot \frac{\log e}{\log \ell} + \ldots + \frac{\log e}{\log^{(r)}(\ell)} \cdot \frac{\log e}{\log^{(r-1)}(\ell)} \ldots \frac{\log e}{\log \ell} < 3$

(C19)

It is quite clear that each term in the L.H.S. of (C19) is bounded above by $\frac{(\log e)^2}{\log \ell}$ and there are $r$ such terms. So it is sufficient to prove that

$$\frac{(\log e)^2}{\log \ell} \cdot r < 3 \text{ for } r \geq 2$$

(C20)

But $\ell \geq \exp_2^{(r)}(2)$; hence it is sufficient to prove that

$$\frac{(\log e)^2}{\exp_2^{(r-1)}(2)} \cdot r < 3 \text{ for } r \geq 2$$

(C21)

which is obviously true.
Thus we have shown that for $\ell \geq 4$, the slope of $f(\ell)$ is bounded by the slope of $a(\ell)$. But we know that $a(\ell)$ increases by 2 when $\ell$ is multiplied by a factor of $\sqrt{2}$. Therefore $f(\ell), \ell \geq 4$ increases by at most 2 every time $\ell$ is multiplied by $\sqrt{2}$. It is trivial to see that for $1 \leq \ell \leq 4$ the difference between $f(\ell)$ and $F^*(\ell)$ cannot exceed 2. This completes the proof of (C15).
References


