Competitive Advantage of Huffman and Shannon-Fano Codes *

Spencer Congero and Kenneth Zeger

IEEE Transactions on Information Theory
Submitted: November 13, 2023

Abstract

For any finite discrete source, the competitive advantage of prefix code C_1 over prefix code C_2 is the probability C_1 produces a shorter codeword than C_2, minus the probability C_2 produces a shorter codeword than C_1. For any source, a prefix code is competitively optimal if it has a nonnegative competitive advantage over all other prefix codes. In 1991, Cover proved that Huffman codes are competitively optimal for all dyadic sources. We prove the following asymptotic converse: As the source size grows, the probability a Huffman code for a randomly chosen non-dyadic source is competitively optimal converges to zero. We also prove: (i) For any source, competitively optimal codes cannot exist unless a Huffman code is competitively optimal; (ii) For any non-dyadic source, a Huffman code has a positive competitive advantage over a Shannon-Fano code; (iii) For any source, the competitive advantage of any prefix code over a Huffman code is strictly less than $\frac{1}{3}$; (iv) For each integer $n > 3$, there exists a source of size n and some prefix code whose competitive advantage over a Huffman code is arbitrarily close to $\frac{1}{3}$; and (v) For each positive integer n, there exists a source of size n and some prefix code whose competitive advantage over a Shannon-Fano code becomes arbitrarily close to 1 as $n \longrightarrow \infty$.

* **S. Congero and K. Zeger** are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 (scongero@ucsd.edu and ken@zeger.us).

[*The Table of Contents below is provided for the reader's convenience and will be removed later*.]

Contents

1. Introduction

In a probabilistic game where multiple players are each rated by a numerical score, one way to designate a particular player A as being superior among their fellow competitors is by "expected score optimality", where no other player B can obtain a better score on average than the score of A. A second way, called "competitive optimality", occurs if for every other player B, the probability of A scoring better than B is at least the probability of B scoring better than A. That is, A has a nonnegative "competitive advantage"

$$P(A \text{ scores better than } B) - P(B \text{ scores better than } A)$$

over all other players B (where tie scores are ignored). In this paper we obtain results about competitive advantage and optimality when lossless source coding is viewed as a game, source codes are the players, the numerical score is the length of the codeword of a randomly chosen source symbol, and a codeword length is deemed to be better than another if its length is shorter. We first formalize some terminology and definitions and then explain the history of the problem and our results.

An *alphabet* is a finite set S, and a *source* of size n with alphabet S is a random variable X such that $|S| = n$ and $P(X = y) = P(y)$ for all $y \in S$. We denote the probability of any subset $B \subseteq S$ by $P(B) = \sum_{y \in B} P(y)$. A source is said to be *dyadic* if $P(y)$ is a nonnegative integer power of $1/2$ for all $y \in S$.

A *code* for a given source X is a mapping $C : S \longrightarrow \{0,1\}^*$ and the binary strings $C(1), \dots, C(n)$ are called *codewords* of C. A *prefix code* is a code where no codeword is a prefix of any other codeword.

A *code tree* for a prefix code C is a rooted binary tree whose leaves correspond to the codewords of C; specifically, the codeword associated with each leaf is the binary word denoting the path from the root to the leaf. The *length* of a code tree node is its path length from the root. The rth *row* of a code tree is the set of nodes whose length is r, and we will view a code tree's root as being on the top of the tree with the tree growing downward. For example, row r of a code tree is "higher" in the tree than row $r + 1$. If x and y are nodes in a code tree, then x is a *descendent* of y if there is a downward path of length zero or more from y to x. Two nodes in a tree are called *siblings* if they have the same parent. In a code tree, for any collection A of nodes having no common leaf descendents, define $P(A)$ to be the probability of the set of all leaf descendents of A in the tree.

A (binary) *Huffman tree* is a code tree constructed from a source by recursively combining two smallest-probability nodes until only one node with probability 1 remains. The initial source probabilities correspond to leaf nodes in the tree. A *Huffman code* for a given source is a mapping of source symbols to binary words by assigning the source symbol corresponding to each leaf in the Huffman tree to the binary word describing the path from the root to that leaf. A *Shannon-Fano code* is a prefix code, such that for each $y \in S$ the codeword associated with the source symbol y has length $\lceil \log_2 \frac{1}{P(y)} \rceil$.

Given a source with alphabet S and a prefix code C, for each $y \in S$ the length of the binary codeword $C(y)$ is denoted $l_C(y)$. Two codes C_1 and C_2 are *length equivalent* if $l_{C_1}(y) = l_{C_2}(y)$ for every source symbol $y \in S$. The *average length* of a code C for a source with alphabet S is $\sum_{y \in S} l_C(y)P(y)$. A prefix code is *expected length optimal* for a given source if no other prefix code achieves a smaller average codeword length for the source.

A code is *complete* if every non-root node in its code tree has a sibling, or, equivalently, if every node has either zero or two children. A code C for a given source is *monotone* if for any two nodes in the code tree of C, we have $P(u) \geq P(v)$ whenever $l_C(u) < l_C(v)$. Expected length optimal codes are always monotone (see Lemma 2.3).

Huffman codes are known to be expected length optimal, monotone, and complete for every source (e.g., see [7]), whereas Shannon-Fano codes are monotone, but need not be expected length optimal nor complete. Shannon-Fano codes are known to achieve the same average lengths as Huffman codes whenever the source is dyadic. For non-dyadic sources, Shannon-Fano codes always have larger average lengths than Huffman codes, but nevertheless the average length of the Shannon-Fano code (and thereby also the Huffman code) is less than one bit larger than the source entropy [7].

The *Kraft sum* of a sequence of nonnegative integers $l_1, \ldots, l_k$ is $2^{-l_1} + \cdots + 2^{-l_k}$. We extend the definition of "Kraft sum" to also apply to sets of source symbols or sets of nodes in a code tree as follows. The Kraft sum $K(A)$ of any collection A of leaves is the Kraft sum of the corresponding sequence of codeword lengths. The Kraft sum $K(A)$ of any collection A of nodes having no common leaf descendents is the Kraft sum of the set of all leaf descendents of A in the tree. The Kraft sum of a collection of source symbols is the Kraft sum of the corresponding leaves in a code tree. The well-known Kraft inequality and its converse are stated next.

Lemma 1.1 (Kraft, e.g., [7, Theorem 5.2.1])**.** *The codeword lengths $l_1, \ldots, l_n$ of any prefix code satisfy $2^{-l_1}+\cdots+2^{-l_n} \leq 1$. Conversely, if a sequence $l_1, \ldots, l_n$ of positive integers satisfies $2^{-l_1}+\cdots+2^{-l_n} \leq 1$, then there exists a binary prefix code whose codeword lengths are $l_1, \ldots, l_n$.*

The following lemma[1] expresses an equality condition for Lemma 1.1.

Lemma 1.2. *A prefix code is complete if and only if its Kraft sum equals* 1*.*

We have defined sources and prefix codes to be finite throughout this paper, but we note that a complete infinite prefix code need not have Kraft sum 1 (e.g., see [12]).

For a subset A of a source's alphabet and for a Huffman code H for the source, we say the *Huffman-Kraft sum* of A is

$$K(A) = \sum_{x \in A} 2^{-l_H(x)},$$

where we reuse the "K" notation. Note that the Huffman-Kraft sum of A is the (usual) Kraft sum of the Huffman codeword lengths of the symbols in A.

As previously mentioned, one measure of the success of a source code that generally differs from expected length optimality is called "competitive optimality" and has been considered in the form of one-on-one competitions between pairs of prefix codes to determine in each competition which code has the higher probability of producing a shorter codeword for a given source.

For a given source with alphabet S, and for any codes C_1 and C_2, define

$$\begin{aligned} W &= \{i \in S : l_{C_1}(i) < l_{C_2}(i)\} \\ L &= \{i \in S : l_{C_1}(i) > l_{C_2}(i)\} \\ T &= \{i \in S : l_{C_1}(i) = l_{C_2}(i)\}. \end{aligned} \tag{1}$$

The sets W, L, and T contain the source values of the *wins*, *losses*, and *ties*, respectively, for code C_1 in a one-on-one competition against code C_2 to see which produces shorter length codewords for the given source. Even though the notation for W, L, and T does not explicitly reference C_1 and C_2, the codes involved will be clear from context. Code C_1 is said to *competitively dominate* code C_2 if $P(W) \geq P(L)$,

[1]The proof follows easily from the proof of Theorem 5.1.1 in the Cover-Thomas textbook [7]. A more general result can be found in Theorem 2.5.19 of the Berstel-Perrin-Reutenauer textbook [2].

and *strictly competitively dominate* code C_2 if $P(W) > P(L)$. A prefix code is *competitively optimal* for a source of size n if it competitively dominates all other prefix codes for the same source. The notion of competitive optimality dates back at least to 1980 in the field of financial investment [1].

In 1991, Cover proved that Shannon-Fano codes are competitively optimal for dyadic sources. Since Huffman and Shannon-Fano codes are length equivalent for dyadic sources (via [7, Theorem 5.3.1]), Huffman codes are also competitively optimal in this case, as reworded below.

Theorem 1.3 (Cover [6, Theorem 2])**.** *Huffman codes are competitively optimal for all dyadic sources.*

Cover's proof also showed that for all dyadic sources the Huffman code strictly competitively dominates all other (i.e., not length equivalent) prefix codes. Additionally, he showed that for all non-dyadic sources Shannon-Fano codes competitively dominate all other prefix codes if an extra one-bit penalty is assessed to the non-Shannon-Fano code during each symbol encoding. This comparison favorably treats ties as if they were wins for the Shannon-Fano code and treats one-bit losses for the Shannon-Fano code as if they were ties, thus giving the Shannon-Fano code a one-bit handicap in betting parlance. As dyadic sources are rare among all sources, it is natural to ask whether Cover's (non-handicapped) competitive optimality result extends in some way to non-dyadic sources.

In 1992, Feder [9] showed that for all non-dyadic sources Huffman codes competitively dominate all other prefix codes if an extra one-bit penalty is assessed to the non-Huffman code during each symbol encoding. This result is analogous to (but uses a different proof technique from) Cover's result for Shannon-Fano codes.

In 1995, Yamamoto and Itoh [16] illustrated a non-dyadic source whose Huffman code was not competitively optimal. They presented a prefix code for the source which strictly competitively dominates the Huffman code, and whose win and loss probabilities are $P(W) = 0.5$ and $P(L) = 0.4$, respectively. Their example consists of a source of size 4 and distinct prefix codes C_1, C_2, and C_3, such that C_i competitively dominates C_j whenever $(i, j) \in \{(1, 2), (2, 3), (3, 1)\}$. This example demonstrates that the relation of competitive dominance is not transitive. One of these codes was the Huffman code, so the Huffman code is not competitively optimal, and no competitively optimal code exists in this case. They also provided a sufficient condition for a source not to have a competitively optimal Huffman code.

Their results, however, do not provide an indication of how many or few source codes would have competitively optimal Huffman codes.

In 2001, Yamamoto and Yokoo [17] studied competitive optimality for almost instantaneous variable-to-fixed length codes.

In 2021, Bhatnagar [3] studied the use of competitive optimality for analyzing error probabilities in artificial intelligence systems.

The Cover-Thomas textbook [7, p. 131] gives the following gambling interpretation of competitive optimality and acknowledges the difficulty of mathematically analyzing Huffman codes:

> "To formalize the question of competitive optimality, consider the following two-person zero-sum game: Two people are given a probability distribution and are asked to design an instantaneous code for the distribution. Then a source symbol is drawn from this distribution, and the payoff to player A is 1 or -1, depending on whether the codeword of player A is shorter or longer than the codeword of player B. The payoff is 0 for ties.
>
> Dealing with Huffman code lengths is difficult, since there is no explicit expression for the codeword lengths."

To determine whether a code is competitively optimal, one must determine if the difference between the probabilities of winning and losing a competition for shortest codeword length is nonnegative for every possible opponent code. When a code C_1 is determined not to be competitively optimal, at least one other code C_2 has a larger probability of winning against C_1 than C_1 has against C_2. It has not previously been known how large or small the difference between those probabilities can be when, say, C_1 is a Huffman or Shannon-Fano code. We define terminology for that probability difference and then proceed to derive an upper bound for it, and describe the tightness of the bound.

For a given source, the *competitive advantage* of code C_1 over code C_2 is the quantity

$$\Delta = P(W) - P(L).$$

Thus, $|\Delta| \leq 1$, and C_1 competitively dominates C_2 if and only if $\Delta \geq 0$. Whereas competitive optimality indicates whether one code always dominates other codes, competitive advantage quantifies by how much one code dominates over another code. If the codes C_1 and C_2 are complete, then their Kraft sums each equal 1 by Lemma 1.2; if additionally C_1 and C_2 are monotone and differ in codeword lengths for at least one source symbol, then it is impossible for C_1 to have codewords that are shorter than or equal in length to those of C_2 for every source symbol, so $\Delta \neq 1$.

The example in [16] gives a non-competitively-optimal Huffman code with a competitive advantage of $0.5 - 0.4 = 0.1$ over a Huffman code. Under the game-playing description of source coding in [7] quoted above, the competitive advantage of Player A's code over Player B's code is equal to the total expected earnings of Player A.

Example 1.4 (Golf analogy)**.** *One round of professional golf typically consists of a collection of players competing for* 18 *holes. Each player's score for a hole is the number of strokes it takes that player to get the ball in the hole. A player's total score for the round is the sum of the player's scores for all* 18 *holes. The player with the lowest total score for the* 18 *holes is the winner of that round. This scoring method is sometimes called "stroke play" or "medal play". Another form of golf scoring is known as "match play", where (perhaps two) players compete against each other, and each hole results in either a "win", a "loss", or a "tie" depending on which player had the lowest number of strokes for that hole. A player's total score for the round is then the total number of wins that player achieved. A player wins the round if the player has more wins than the other player. Stroke play and match play golf scoring are analogous to average codeword length and competitive advantage, respectively, in lossless coding, if one associates the golf players with prefix codes, golf holes with uniformly drawn source symbols, and player scores for a hole with codeword lengths produced by a prefix code.*

The following questions have not previously been answered in the literature: (i) How likely or unlikely is it for a Huffman code to be competitively optimal, allowing for non-dyadic sources?; (ii) If a Huffman code is not competitively optimal for a particular source, how large can the competitive advantage Δ of another code be over the Huffman code?; (iii) Do Huffman codes always (perhaps, strictly) competitively dominate Shannon-Fano codes for non-dyadic sources, and how large can the competitive advantage Δ of another code be over the Shannon-Fano code?

One approach we exploit to answer the first question is to choose a source "at random" and seek the probability that a resulting Huffman code is competitively optimal. For the second and third questions, one can seek the best possible upper bound on the competitive advantage over a Huffman code or over a Shannon-Fano code.

In this paper we address these questions by proving the following main results: (1) Competitively optimal codes can exist for a given source only if some Huffman code is competitively optimal for that source (Theorem 2.7); (2) The probability that a Huffman code for a rather generally chosen random source is competitively optimal converges to zero as the source size grows (Theorem 3.5), and therefore the probability that competitively optimal codes exist for such sources also converges to zero (Corollary 3.7); (3) For all non-dyadic sources, Huffman codes strictly competitively dominate Shannon-Fano codes (Theorem 5.1); (4) For all non-dyadic sources, the competitive advantage Δ of any code over a Huffman code is strictly less than $\frac{1}{3}$ (Theorem 6.6); (5) For each integer $n > 3$, there exists a non-dyadic source of size n and some prefix code whose competitive advantage Δ over a Huffman code is arbitrarily close to $\frac{1}{3}$ (Theorem 6.7); (6) For each positive integer n, there exists a non-dyadic source of size n and a prefix code for the source such that the competitive advantage Δ of the code over a Shannon-Fano code for the source becomes arbitrarily close to 1 as $n \longrightarrow \infty$, and the average length of the code becomes arbitrarily close to one bit less than the average length of the Shannon-Fano code as $n \longrightarrow \infty$ (Theorem 7.1).

We also analyze "small" sources and show that for all sources of size at most 3, Huffman codes are competitively optimal (Theorem 8.1), and sources of size 4 for which Huffman codes are competitively optimal can be characterized in terms of a certain convex polyhedral condition (Theorem 8.2).

Finally we conducted computer simulations that drew a million random sources from a flat Dirichlet distribution for each source size up to 34 source symbols and determined whether the resulting Huffman code satisfies a sufficient condition for competitively optimality. Our numerical observations are given in Section 9 and indicate that the convergence proven in Theorem 3.5 is very rapid for relatively small source sizes (Figure 2).

2. Existence of competitively optimal codes

In this section, we use a result of Yamamoto and Itoh to show that competitively optimal codes can exist for a given source only if some Huffman code is competitively optimal for that source (Theorem 2.7). This result is later used in Section 3 to show that competitively optimal codes usually do not exist for large randomly chosen sources (Corollary 3.7).

If a source is dyadic, then there can be only one Huffman code (up to length equivalence) and this code is length equivalent to a Shannon-Fano code. However, for non-dyadic sources, somewhat unusual circumstances can arise. In what follows, Example 2.1 illustrates a source where one Huffman code is competitively optimal but another Huffman code for the same source is not, and Example 2.2 illustrates a source with two Huffman codes and a non-Huffman code that form a cycle of strict competitive domination, as well as another non-Huffman code that is expected length optimal.

Example 2.1 (Two Huffman codes)**.**
A source with symbols a, b, c, d and corresponding probabilities $\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}$ has a Huffman code H_1 with codeword lengths $2, 2, 2, 2,$ and another Huffman code H_2 with lengths $1, 2, 3, 3.$ Each Huffman code has competitive advantage zero over the other. However, one can verify that H_1 is competitively optimal, whereas H_2 is not.

Example 2.2 (Two Huffman codes and two other codes)**.**
A source with symbols a, b, c, d, e, f and corresponding probabilities $\frac{1}{3}, \frac{1}{3}, \frac{1}{9}, \frac{1}{9}, \frac{1}{18}, \frac{1}{18}$ has a Huffman code H_1 with codeword lengths $1, 2, 3, 4, 5, 5$ and another Huffman code H_2 with lengths $2, 2, 3, 3, 3, 3$ (see Figure 1), each with average codeword length equal to $\frac{7}{3}$. Two other trees for codes C_1 and C_2 are shown, which are relabeled versions of the trees for H_2 and H_1, respectively. H_1 strictly competitively dominates H_2 since its competitive advantage is $P(a) - P(d, e, f) = \frac{1}{9}$, while C_1 strictly competitively dominates H_1 since its competitive advantage is $P(b, c) - P(a) = \frac{1}{9}$, and H_2 strictly competitively dominates C_1 since its competitive advantage is $P(a, d, e, f) - P(b, c) = \frac{1}{9}$. That is, H_1, H_2, and C_1 form a cycle, which illustrates the non-transitivity of strict competitive dominance. Also, C_2 is an expected length optimal code, but it is non-Huffman since nodes e and f were not merged.

As observed in the example, although Huffman codes are always expected length optimal for any given source, expected length optimal prefix codes need not be Huffman codes. But it turns out that any expected length optimal code is length equivalent to some Huffman code for the source.

Lemma 2.3 (e.g., [10, p. 670])**.** *For any source, if a prefix code is expected length optimal, then it is monotone.*

The following lemma may be hinted at in the proof of [7, Lemma 5.8.1, p. 123], and is also a special case of Lemma 4.3 in Section 4.

Lemma 2.4 ([5])**.** *For any source, every expected length optimal prefix code is length equivalent to some Huffman code.*

Lemma 2.5 (Yamamoto and Itoh [16, Theorem 3])**.** *For any source, every competitively optimal code is expected length optimal.*

The following corollary follows immediately from Lemma 2.5 and Lemma 4.3 (in Section 4).

Corollary 2.6. *For any source, every competitively optimal code is length equivalent to some Huffman code.*

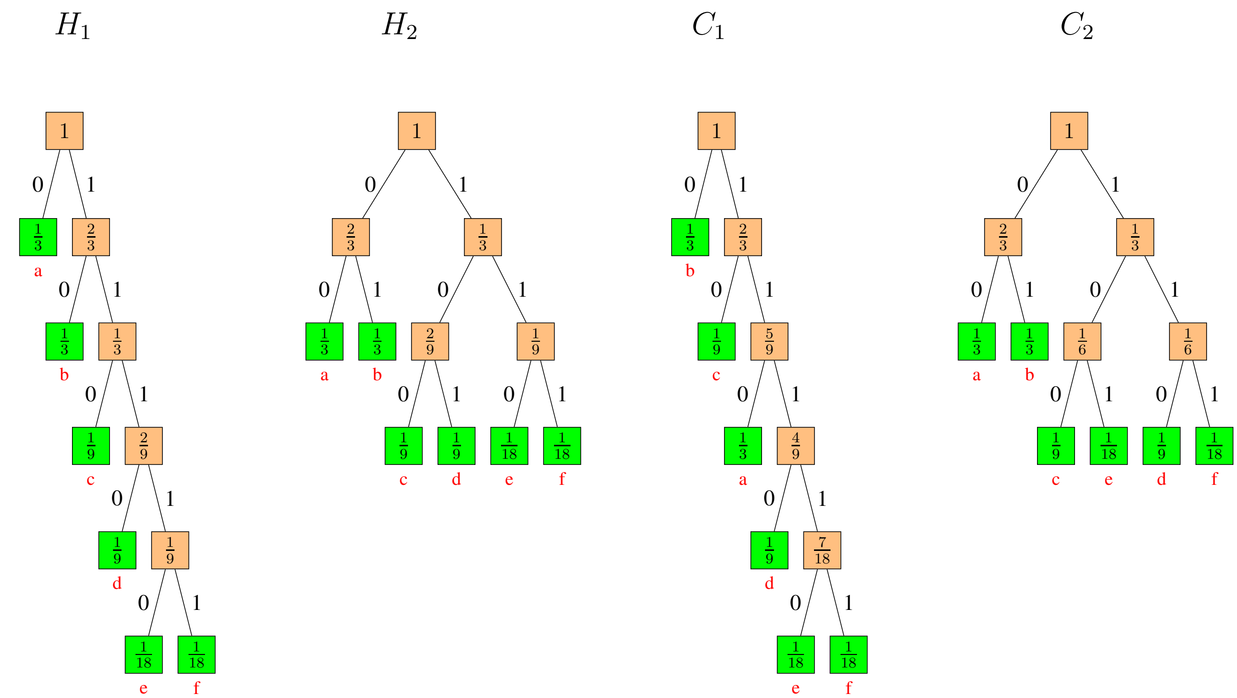

Figure 1: Code trees of four prefix codes for a source of size 6.

Note that the premises in Lemma 2.5 and Corollary 2.6 might be vacuous if no competitively optimal code exists for a particular source, in which case no conclusion can be drawn.

The following theorem shows that competitively optimal codes for a given source can exist only if some Huffman code is competitively optimal for the source.

Theorem 2.7. *For any source, if no Huffman code is competitively optimal, then no prefix code is competitively optimal.*

Proof. To prove the contrapositive, suppose some prefix code C is competitively optimal for a given source. Corollary 2.6 implies that some Huffman code maps source symbols to codewords of the same length as the codewords assigned by C. Thus, for any prefix code C', the competitive advantage of C over C' is nonnegative and is equal to the competitive advantage of the Huffman code over C', so the Huffman code also dominates C'. Thus, the Huffman code is competitively optimal. ■

3. Asymptotic converse to Cover's theorem on competitive optimality of Huffman codes

In this section, our goal is to analyze how likely it is that a source will have a competitively optimal Huffman code (and more generally any competitively optimal code). To address this question, we need to make precise what "how likely" means in this context. There are many possible ways to choose a source "at random". By some means we wish to randomly obtain numbers $p_1, \dots, p_n \in (0, 1)$ whose sum equals 1, interpret them as probabilities, and then determine whether a Huffman code constructed from these probabilities is competitively optimal.

One commonly used way to randomly obtain such probabilities is to sample from a "flat Dirichlet distribution". A *flat Dirichlet distribution* of size n has a uniform probability density on the $(n-1)$-dimensional simplex $\{(x_1, \dots, x_n) \in [0, 1]^n : x_1 + \dots + x_n = 1\}$ embedded in $\mathbb{R}^n$ (e.g., see [14]). For example, when $n = 3$, the point (p_1, p_2, p_3) is chosen uniformly from the (2-dimensional) triangle embedded in $\mathbb{R}^3$, whose vertices are $(0, 0, 1)$, $(0, 1, 0)$, and $(1, 0, 0)$. Choosing a random source from a flat Dirichlet distribution tends to be a natural approach since it treats all coordinates equally and uniformly. This method of random source selection was used, for example, to analyze average Huffman code rate redundancy in [11] and [15].

Another way to randomly create a source of size n is to choose n positive i.i.d. samples $X_1, \dots, X_n$ according to some probability distribution, form their sum $S_n = X_1 + \dots + X_n$, and then construct the normalized sequence $\frac{X_1}{S_n}, \dots, \frac{X_n}{S_n}$. This technique specializes to a flat Dirichlet distribution when $X_1, \dots, X_n$ are i.i.d. exponentials with mean one, as given in the following lemma.

Lemma 3.1 (e.g., [8, Chapter 11, Theorem 4.1])**.** *If $X_1, \dots, X_n$ are i.i.d. exponential random variables with mean one and $S_n = X_1 + \dots + X_n$, then the joint distribution of $\frac{X_1}{S_n}, \dots, \frac{X_n}{S_n}$ is the same as that of a flat Dirichlet distribution of size n.*

Using this more general method for randomly generating a source of size n, we prove in this section that if the density of the i.i.d. sequence is positive on at least some interval $(0, \epsilon)$ with $\epsilon > 0$, then the probability a Huffman code is competitively optimal shrinks to 0 as n grows (Theorem 3.5). In other words, competitively optimal codes become rare for large sources. This result can be viewed as an asymptotic converse to Cover's theorem for dyadic sources (i.e., Theorem 1.3). It also indicates that Cover's result cannot be extended to many large sources beyond the dyadic ones.

We also examine an important special case of Theorem 3.5 when the random variables $X_1, \dots, X_n$ are i.i.d. exponential.

Our Corollary 3.6 notes that the result of Theorem 3.5 is true when choosing a source at random from a flat Dirichlet distribution, which we use in Section 9 to gather experimental evidence of the convergence rate as $n \longrightarrow \infty$.

The next lemma shows that if one set of leaves of a Huffman code has a smaller Kraft sum than that of a second disjoint set of leaves, then one can find a prefix code that competitively wins against the Huffman code on the set of smaller Kraft sum, loses on the set of larger Kraft sum, and ties on all other leaves.

Lemma 3.2. *For any source, if H is a Huffman code, and U and V are disjoint subsets of the source's alphabet S whose Huffman-Kraft sums satisfy $K(U) < K(V)$, then there exists a prefix code C such that*

$$U = \{x \in S : l_C(x) < l_H(x)\}$$
$$V = \{x \in S : l_C(x) > l_H(x)\}.$$

Proof. Let k be an integer such that

$$K(U) \leq (1 - 2^{-k})K(V) \tag{2}$$

and for each $x \in S$, define the following integer:

$$l(x) = \begin{cases} l_H(x) - 1 & \text{if } x \in U \\ l_H(x) + k & \text{if } x \in V \\ l_H(x) & \text{if } x \notin U \cup V. \end{cases}$$

Then

$$\begin{aligned} \sum_{x \in S} 2^{-l(x)} &= \sum_{x \in U} 2 \cdot 2^{-l_H(x)} + \sum_{x \in V} 2^{-k} \cdot 2^{-l_H(x)} + \sum_{x \notin U \cup V} 2^{-l_H(x)} \\ &= 2K(U) + 2^{-k}K(V) + (1 - K(U) - K(V)) \qquad (3) \\ &= 1 + K(U) - (1 - 2^{-k})K(V) \\ &\leq 1, \qquad (4) \end{aligned}$$

where (3) follows since the Kraft sum of all the Huffman codewords is 1, by Lemma 1.2; and (4) follows from (2). Therefore, the Kraft inequality (Lemma 1.1) implies that there exists a prefix code C whose codeword lengths are the values of $l(x)$ for all $x \in S$. That is, $l_C(x) = l_H(x) - 1$ for all $x \in U$; $l_C(x) = l_H(x) + k$ for all $x \in V$; and $l_C(x) = l_H(x)$ for all $x \notin U \cup V$, and therefore

$$\begin{aligned} U &= \{x \in S : l_C(x) < l_H(x)\} \\ V &= \{x \in S : l_C(x) > l_H(x)\}. \end{aligned}$$

■

Lemma 3.3. *For any source and any Huffman code for the source, if U and V are disjoint subsets of the source alphabet whose Huffman-Kraft sums satisfy $K(U) < K(V)$ and whose probabilities satisfy $P(U) > P(V)$, then the Huffman code is not competitively optimal for the source.*

Proof. Let S denote the source alphabet and let H be a Huffman code. By Lemma 3.2, since $K(U) < K(V)$, there exists a prefix code C such that

$$\begin{aligned} U &= \{x \in S : l_C(x) < l_H(x)\} \\ V &= \{x \in S : l_C(x) > l_H(x)\}. \end{aligned}$$

The competitive advantage of C over the Huffman code H is thus $P(U) - P(V) > 0$, so the Huffman code is not competitively optimal. ■

The following lemma is essentially given in [16, equation (4)], but we give an alternate short proof here for completeness.

Lemma 3.4. *For any source, if y and y' are sibling nodes in a Huffman tree, z is a leaf node descendant of y, and $P(z) < P(y) - P(y')$, then the Huffman code is not competitively optimal for the source.*

Proof. We may assume $P(y) > P(y')$ since $P(z) > 0$. Let U and V be the sets of leaf node descendants of y and y', respectively. Let $U' = U - \{z\}$. Then, $K(U') < K(U) = K(V)$ (since siblings have the same

Kraft sum) and $P(U') = P(U) - P(z) > P(U) - (P(U) - P(V)) = P(V)$. Then by Lemma 3.3, the Huffman code is not competitively optimal. ■

A hypothetical converse to Theorem 1.3 would say that Huffman codes for non-dyadic sources are never competitively optimal. This is not true (e.g., Theorem 8.1), but we are able to demonstrate that such a converse becomes probabilistically true as the source size grows.

Our next theorem, one of our main results, demonstrates an asymptotic converse to Theorem 1.3, by showing that competitively optimal Huffman codes become rare for randomly chosen sources as their size grows.

Theorem 3.5. *Let $\epsilon > 0$ and let $X_1, X_2, \ldots$ be an i.i.d. sequence of nonnegative random variables with a density which is positive on at least $(0, \epsilon)$, and let $S_n = X_1 + \cdots + X_n$. Then the probability that a Huffman code for $\frac{X_1}{S_n}, \ldots, \frac{X_n}{S_n}$ is competitively optimal converges to zero as $n \longrightarrow \infty$.*

Proof. Let F denote the distribution function for each X_i. Let $\delta = \epsilon/24$. For each $k \in \{1, \ldots, 24\}$, define the interval

$$I_k = (\epsilon - k\delta, \epsilon - (k-1)\delta).$$

The intervals $I_1, \ldots, I_{24}$ are disjoint and their union lies in $(0, \epsilon)$.

Denote the indicator function for any $E \subseteq \mathbb{R}$ by $\mathbb{1}_E(x) = 1$ if $x \in E$ and $\mathbb{1}_E(x) = 0$ otherwise. For each $k \in \{1, \ldots, 24\}$, since the binary random variables $\mathbb{1}_{I_k}(X_i)$ are i.i.d. for all i, we have

$$\begin{aligned}
\frac{|\{i \in \{1, \ldots, n\} : X_i \in I_k\}|}{n} &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{I_k}(X_i) \\
&\xrightarrow{a.s.} E[\mathbb{1}_{I_k}(X_1)] &\quad (5) \\
&= P(X_1 \in I_k) \\
&= F(\epsilon - (k-1)\delta) - F(\epsilon - k\delta) \\
&> 0, &\quad (6)
\end{aligned}$$

where (5) follows from the strong law of large numbers (e.g., [4, Theorem 6.1]); and (6) follows since F is increasing on $(0, \epsilon)$. Let A be the event that all 24 of the convergences in (5) occur. The intersection of finitely many events with probability 1 has probability 1, so $P(A) = 1$.

We will show that for any outcome $\omega \in A$, there exists $N \geq 1$ such that for all $n \geq N$, any Huffman code for $\frac{X_1(\omega)}{S_n(\omega)}, \ldots, \frac{X_n(\omega)}{S_n(\omega)}$ is not competitively optimal. Suppose this has been done, and for each $n \geq 1$ let B_n be the event containing the outcomes ω such that a Huffman code for $\frac{X_1(\omega)}{S_n(\omega)}, \ldots, \frac{X_n(\omega)}{S_n(\omega)}$ is competitively optimal. If $\omega \in A$, then ω appears in only finitely many B_n, so $\omega \notin \limsup_{n\to\infty} B_n = \bigcap_{n\geq 1} \bigcup_{j\geq n} B_j$. Therefore $\limsup_{n\to\infty} B_n \subseteq A^c$, and so the theorem is proved using

$$\limsup_{n\to\infty} P(B_n) \leq P(\limsup_{n\to\infty} B_n) \leq P(A^c) = 0$$

where the first inequality follows from Fatou's Lemma (e.g., [4, Theorem 4.1]).

We will now prove what we promised. Let $\omega \in A$, and consider the sequence $X_1(\omega), X_2(\omega), \ldots$. Let H be a Huffman tree constructed from the n probabilities $\frac{X_1(\omega)}{S_n(\omega)}, \ldots, \frac{X_n(\omega)}{S_n(\omega)}$. Since $\omega \in A$, the convergence in (5) holds for all $k \in \{1, \dots, 24\}$, so for each such k there exists $N_k \geq 1$ such that for all $n \geq N_k$ we have $|\{i \in \{1, \dots, n\} : X_i(\omega) \in I_k\}| \geq 1$.

Let $n \geq \max(N_1, \dots, N_{24})$. Then for each $k \in \{1, \dots, 24\}$, let Y_k equal $X_i(\omega)/S_n(\omega)$ for some $X_i(\omega) \in I_k$. That is, $Y_1 > Y_2 > \cdots > Y_{24}$ are probabilities corresponding to 24 of the n leaves in the Huffman tree H. Note that for all k, we have $Y_k S_n(\omega) \in I_k$, so

$$\frac{\epsilon - k\delta}{S_n(\omega)} < Y_k < \frac{\epsilon - (k-1)\delta}{S_n(\omega)}. \tag{7}$$

The sibling in H of the leaf for Y_{14} has probability at most Y_{13}, for otherwise the leaf for Y_{14} and its sibling would not have been nodes with two of the smallest available probabilities for merging as required by the Huffman construction. Then the probability $\widehat{Y}_{14}$ of the parent of the leaf for Y_{14} satisfies (using (7))

$$\widehat{Y}_{14} \leq Y_{14} + Y_{13} < \frac{\epsilon - 13\delta}{S_n(\omega)} + \frac{\epsilon - 12\delta}{S_n(\omega)} = \frac{\epsilon - \delta}{S_n(\omega)} < Y_1.$$

Then since the Huffman code H is monotone by Lemma 2.3, the leaf for Y_1 appears on a row in H that is at least as high as the row on which the parent of Y_{14} appears, so the leaf for Y_1 appears on a row strictly higher than the row on which the leaf for Y_{14} appears. Since $Y_1 > Y_2 > \cdots > Y_{14}$, monotonicity of Huffman codes shows the row numbers on which the leaves for these 14 probabilities appear are non-decreasing, so the conclusion from the previous sentence implies the leaf for some probability in the sequence $Y_1, Y_2, \dots, Y_{14}$ appears on a row strictly higher in H than the leaf for the next probability in the sequence. Specifically, there exists $m \in \{1, \dots, 13\}$ such that the leaf for Y_m appears on a row r strictly higher than the row r' on which the leaf for Y_{m+1} appears, i.e., $r < r'$.

Similarly, the sibling for the leaf for Y_{21} has probability at most Y_{20}, so the probability $\widehat{Y}_{21}$ of the parent of the leaf for Y_{21} satisfies (using (7))

$$\widehat{Y}_{21} \leq Y_{21} + Y_{20} < \frac{\epsilon - 20\delta}{S_n(\omega)} + \frac{\epsilon - 19\delta}{S_n(\omega)} = \frac{\epsilon - 15\delta}{S_n(\omega)} < Y_{14} \leq Y_{m+1}. \tag{8}$$

Then since Huffman codes are monotone, the leaf for Y_{m+1} appears on a row that is at least as high as the row on which the parent of Y_{21} appears, so the row r' on which the leaf for Y_{m+1} appears is strictly higher than the row r'' on which the leaf for Y_{21} appears, i.e., $r' < r''$.

Let U be the set consisting of the leaves for Y_{m+1} and Y_{21}, and let V be the set consisting of the leaf for Y_m. Then

$$K(U) = 2^{-r'} + 2^{-r''} < 2^{-(r'-1)} \leq 2^{-r} = K(V),$$

and again using (7),

$$P(V) - P(U) = Y_m - (Y_{m+1} + Y_{21}) < \frac{\epsilon - (m-1)\delta}{S_n(\omega)} - \left(\frac{\epsilon - (m+1)\delta}{S_n(\omega)} + \frac{\epsilon - 21\delta}{S_n(\omega)} \right) = \frac{-\delta}{S_n(\omega)} < 0.$$

Then Lemma 3.3 implies the Huffman code for $\frac{X_1(\omega)}{S_n(\omega)}, \dots, \frac{X_n(\omega)}{S_n(\omega)}$ is not competitively optimal, which is what we wanted to show. ∎

The requirements for the density of the random variables chosen in the statement of Theorem 3.5 are not very restrictive and are satisfied by a wide range of random variables. Such densities include various versions of at least the following: beta, chi-square, Erlang, exponential, gamma, Gumbel, log-normal, logistic, Maxwell, Pareto, Rayleigh, uniform, and Weibull.

The following corollary follows from Lemma 3.1 and Theorem 3.5. It shows that competitively optimal Huffman codes become rare for large sources selected uniformly at random from a simplex.

Corollary 3.6. *Let $X_1, \ldots, X_n$ be the probabilities of a source chosen randomly from a flat Dirichlet distribution in $\mathbb{R}^n$. Then the probability that a Huffman code for this source is competitively optimal converges to zero as $n \longrightarrow \infty$.*

The next corollary follows immediately from Theorem 2.7 and Theorem 3.5. It shows that as the source size grows the existence of any competitively optimal codes becomes unlikely.

Corollary 3.7. *Let $\epsilon > 0$ and let $X_1, X_2, \ldots$ be an i.i.d. sequence of nonnegative random variables with a density which is positive on at least $(0, \epsilon)$, and let $S_n = X_1 + \cdots + X_n$. Then the probability that a competitively optimal code exists for the source $\frac{X_1}{S_n}, \ldots, \frac{X_n}{S_n}$ converges to zero as $n \longrightarrow \infty$.*

4. Lemmas for future sections

In this section, we give a number of lemmas that are used in the remainder of this paper.

Huffman codes are expected length optimal and Gallager [10, Theorem 1] characterized such codes in terms of a "sibling property", which says the code is complete and if the nodes in the code tree can be listed in order of non-increasing probability with each node being adjacent in the list to its sibling.

However, Huffman codes are not the only expected length optimal prefix codes. Expected length optimal prefix codes are characterized as those which are length equivalent to some Huffman code for the source. Below we prove a second characterization of expected length optimal prefix codes, or equivalently prefix codes which are length equivalent to a Huffman code. Namely we show that such codes are strongly monotone. Strong monotonicity is a weaker condition than the sibling property, and thus a broader class of prefix codes are strongly monotone.

Let C be a prefix code a source has alphabet S. If $U \subset S$, then denote the Kraft sum of C's codeword lengths corresponding to the source symbols in U by

$$K_C(U) = \sum_{x \in U} 2^{-l_C(x)}.$$

Definition 4.1. Given a source with alphabet S, a prefix code C for the source is *strongly monotone* if for any subsets $A, B \subseteq S$, if there exist integers $i, j \geq 0$ such that $K_C(A) = 2^{-i} > 2^{-j} = K_C(B)$, then $P(A) \geq P(B)$.

The following lemma notes several properties that are preserved under length equivalence between prefix codes.

Lemma 4.2 ([5]). *If two prefix codes are length equivalent, then each of the following properties holds for one code if and only if it holds for the other code:*

- *completeness*
- *strong monotonicity*
- *expected length optimality.*

Lemma 4.3 ([5, Theorem 2.4]). *For any source and any prefix code C for the source, the following are equivalent:*

(1) C is complete and strongly monotone

(2) C is length equivalent to a Huffman code

(3) C is expected length optimal.

For a given source and a Huffman code, if A is a nonempty proper subset of the source alphabet whose Huffman-Kraft sum has binary expansion $K(A) = 0.b_1 b_2 \ldots$, then we define a *Huffman-Kraft partition* of A to be any sequence $A_1, A_2, \ldots$ of disjoint (possible empty) subsets of A whose union is A and such that $K(A_i) = b_i 2^{-i}$ for each i. If $b_k = 1$ and $b_i = 0$ for all $i \geq k+1$, then it suffices to specify the first k sets $A_1, \ldots, A_k$ in a Huffman-Kraft partition.

Lemma 4.4. *Every nonempty proper subset of a source's alphabet has a Huffman-Kraft partition.*

Proof. Let A be a nonempty proper subset of a source's alphabet and let $0.a_1a_2\dots a_k$ be the binary expansion of the Huffman-Kraft sum $K(A)$, where $a_k = 1$.

We use induction on $N = a_1 + \cdots + a_k$. Since A is nonempty, $K(A) > 0$, so $N \geq 1$. First, suppose $N = 1$. Then $a_i = 0$ for all $i \in \{1, \dots, k-1\}$, so $K(A) = 2^{-k}$. Setting $A_k = A$ and $A_i = \varnothing$ for $i \in \{1, \dots, k-1\}$ gives a Huffman-Kraft partition of A.

Now let $m \geq 2$ and suppose the lemma holds for $N = m - 1$. We will next show the lemma must hold for $N = m$. Let $X = \{u \in A : K(u) \leq 2^{-k}\}$. Then $K(A - X)$ is an integer multiple of 2^{-k+1}, since for all $v \in A - X$ we have $K(v) = 2^{-k+i}$ for some $i \geq 1$. Thus, the binary expansion of $K(A - X)$ has 0s in all positions $i \geq k$. Therefore, since $a_k = 1$ and $K(A) = K(X) + K(A - X)$, the binary expansion of $K(X)$ has a 1 in position k, so $K(X) \geq 2^{-k}$.

Let Y be a subset of X whose Huffman-Kraft sum is minimum among all subsets of X with Huffman-Kraft sum at least 2^{-k}. We will show that $K(Y) = 2^{-k}$. Suppose to the contrary that $K(Y) > 2^{-k}$. Let $y \in Y$ be an element with minimum probability among the elements of Y. Since $y \in X$, we know $K(y) \leq 2^{-k}$, and also $K(y) = 2^{-i}$ for some integer i. Therefore, 2^{-k} is an integer multiple of $K(y)$. Also, $K(Y)$ is an integer multiple of $K(y)$, since for all $u \in Y$ there exists an $i \geq 0$ such that $K(u) = 2^i K(y)$. Therefore, $K(Y) \geq 2^{-k} + K(y)$ since $K(Y) > 2^{-k}$. But then $K(Y - \{y\}) = K(Y) - K(y) \geq 2^{-k}$ and $K(Y - \{y\}) = K(Y) - K(y) < K(Y)$, so the Huffman-Kraft sum of $Y - \{y\}$ is at least 2^{-k} but is smaller than that of Y, contradicting the minimality assumption on Y. Therefore, $K(Y) = 2^{-k}$.

Set $A_k = Y$. Since $m \geq 2$, the set $A - A_k$ is nonempty, and since A is a proper subset of the source's alphabet, so is $A - A_k$. Also, the Huffman-Kraft sum $K(A - A_k) = K(A) - K(A_k) = K(A) - 2^{-k}$ has exactly $m - 1$ ones in its binary expansion. Thus, the induction hypothesis implies that $A - A_k$ has a Huffman-Kraft partition $A_1, \dots, A_{k-1}$. Then $A_1, \dots, A_k$ is a Huffman-Kraft partition of A. ■

Lemma 4.5. *Let A and B be subsets of a source alphabet whose Huffman-Kraft sums satisfy $K(A) = K(B) = 2^{-i}$ for some integer $i \geq 0$, and such that $|A| \geq 2$. Then $P(A) \leq 2P(B)$.*

Proof. By Lemma 4.3, Huffman codes are strongly monotone, so Huffman-Kraft sums obey the strong monotonicity condition.

Let $a \in A$ be an element of minimum Huffman-Kraft sum among the elements of A. Since $|A| \geq 2$, there exists $m \geq i + 1$ such that $K(a) = 2^{-m}$, for otherwise $K(A) \geq 2 \cdot 2^{-i} > 2^{-i} = K(A)$, a contradiction. Therefore, the binary expansion of $K(A - \{a\}) = K(A) - K(a) = 2^{-i} - 2^{-m}$ has 1s in positions $i + 1, \dots, m$ and 0s in all other positions. By Lemma 4.4, there exists a Huffman-Kraft partition $A_1, \dots, A_m$ of $A - \{a\}$. In particular, $K(A_{i+1}) = 2^{-(i+1)}$, and so $K(A - A_{i+1}) = K(A) - K(A_{i+1}) = 2^{-i} - 2^{-(i+1)} = 2^{-(i+1)}$. Then by strong monotonicity $P(A_{i+1}), P(A - A_{i+1}) \leq P(B)$, and so $P(A) = P(A_{i+1}) + P(A - A_{i+1}) \leq 2P(B)$. ■

Lemma 4.6. *Given a source with alphabet S, suppose there is a subset $U \subseteq S$ whose Huffman-Kraft sum $K(U)$ is an integer multiple of 2^{-i} for some integer $i \geq 0$. If $A \subseteq U$ with $0 < K(A) < 2^{-j}$ for some integer $j \geq i$, then there exists a subset $B \subseteq U - A$ with $K(A \cup B) = 2^{-j}$.*

Proof. Since $K(A) > 0$ and $K(U - A) = K(U) - K(A) > 2^{-i} - 2^{-j} \geq 0$, Lemma 4.4 shows there exist Huffman-Kraft partitions $A_1, A_2, \dots$ of A, and $B_1, B_2, \dots$ of $U - A$. Let

$$B = \bigcup_{k>j} B_k.$$

Then $K(B) = K(B_{j+1}) + K(B_{j+2}) + \cdots \leq 2^{-(j+1)} + 2^{-(j+2)} + \cdots = 2^{-j}$, so $K(A \cup B) = K(A) + K(B) < 2^{-(j-1)}$. Since $K(U) = K(A \cup B) + K(B_1 \cup \cdots \cup B_j)$, and both $K(U)$ and $K(B_1 \cup \cdots \cup B_j)$

are integer multiples of 2^{-j}, it must be the case that $K(A \cup B)$ is an integer multiple of 2^{-j} as well. Then $0 < K(A \cup B) < 2^{-(j-1)}$ implies $K(A \cup B) = 2^{-j}$. ■

Lemma 4.7. *Given a source with alphabet S, suppose $A, B \subseteq S$ with Huffman-Kraft sums satisfying $2K(B) \leq 2^{-i} \leq K(A)$ for some integer i. Then $P(A) \geq P(B)$, with equality possible only if $K(A) = 2K(B)$.*

Proof. If B is empty, then $P(B) = 0$ and the result follows, so assume B is nonempty.

If $K(B) < 2^{-(i+1)}$, then by Lemma 4.6 (where A, B, U, i, j in Lemma 4.6 respectively correspond to B, B', S, 0, $i+1$ here) there exists a subset $B' \subseteq S - B$ with $K(B') = 2^{-(i+1)} - K(B)$. Alternatively, if $K(B) = 2^{-(i+1)}$ then we will let $B' = \varnothing$. In either case, $K(B \cup B') = 2^{-(i+1)}$.

By Lemma 4.4, there exists a Huffman-Kraft partition $A_1, A_2, \ldots$ of A. Let k be the smallest integer such that $A_k \neq \varnothing$. Since $K(A) \geq 2^{-i}$, we have $k \leq i$.

If $k < i$, then since $0 < K(B \cup B') = 2^{-(i+1)} < 2^{-(k+1)}$, Lemma 4.6 implies (where A, B, U, i, j in Lemma 4.6 respectively correspond to $B \cup B'$, E, S, 0, $i+1$ here) there exists a subset $E \subseteq S - (B \cup B')$ such that $K(E) = 2^{-(k+1)} - K(B \cup B')$. If $k = i$, then instead let $E = \varnothing$. In either case, $K(B \cup B' \cup E) = 2^{-(k+1)}$.

We have

$$\begin{aligned} P(A) &= P(A_k) + P(A - A_k) \\ &\geq P(A_k) && (9) \\ &\geq P(B \cup B' \cup E) && (10) \\ &= P(B) + P(B') + P(E) && (11) \\ &\geq P(B). && (12) \end{aligned}$$

where (10) follows the Huffman-Kraft sum equality $K(A_k) = 2^{-k} = 2K(B \cup B' \cup E)$ since Huffman codes are strongly monotone by Lemma 4.3; and (11) follows since B, B', and E are disjoint.

Now suppose $K(A) > 2K(B)$. Then either $K(B) < 2^{-(i+1)}$ or $K(A) > 2^{-i}$. In the first case, we have $P(B') > 0$ since $K(B \cup B') = 2^{-(i+1)}$, so the inequality in (12) becomes strict. Now consider two subcases of the second case. If $K(A) < 2^{-(i-1)}$, then $i - 1 < k \leq i$, so $k = i$ which implies $K(A - A_k) = K(A - A_i) = K(A) - K(A_i) = K(A) - 2^{-i} > 0$, and thus the inequality in (9) becomes strict. Otherwise, if $K(A) \geq 2^{-(i-1)}$, then $k < i$, and so $K(E) = 2^{-(k+1)} - K(B \cup B') = 2^{-(k+1)} - 2^{-(i+1)} > 0$. Thus $P(E) > 0$, and the inequality in (12) becomes strict. ■

5. Huffman codes competitively dominate Shannon-Fano codes

For dyadic sources, there can be only one Huffman code (up to length equivalence) and such a code is length equivalent to a Shannon-Fano code [7, Theorem 5.3.1], and thus the competitive advantage of either code over the other code is $\Delta = 0$. In this section, Theorem 5.1 shows that for all non-dyadic sources, every Huffman code always strictly competitively dominates every Shannon-Fano code.

This result shows that Shannon-Fano codes are never competitively optimal for non-dyadic sources. Actually, as the source size grows, Huffman codes themselves are usually not competitively optimal either, as shown in Section 3.

The following theorem shows that when a source is not dyadic, every Huffman code always has a positive competitive advantage over the Shannon-Fano code.

Theorem 5.1. *Huffman codes strictly competitively dominate Shannon-Fano codes if and only if the source is not dyadic.*

Proof. Suppose a non-dyadic source has alphabet S and a Huffman code H. By Lemma 4.3, Huffman codes are strongly monotone, so Huffman-Kraft sums obey the strong monotonicity condition.

Denote the Huffman and Shannon-Fano codeword lengths for each $y \in S$ by $l_{\mathrm{H}}(y)$ and $l_{\mathrm{SF}}(y)$, respectively. Let $W = \{i \in S : l_{\mathrm{SF}}(y) < l_{\mathrm{H}}(y)\}$ and $L = \{i \in S : l_{\mathrm{SF}}(y) > l_{\mathrm{H}}(y)\}$ and $T = S - (W \cup L)$, as in (1). It suffices to show $P(W) < P(L)$.

If $y \in W$ then

$$\log_2 \frac{1}{P(y)} \leq \left\lceil \log_2 \frac{1}{P(y)} \right\rceil = l_{\mathrm{SF}}(y) \leq l_{\mathrm{H}}(y) - 1,$$

and if $y \in L$ then

$$\log_2 \frac{1}{P(y)} > \left\lceil \log_2 \frac{1}{P(y)} \right\rceil - 1 = l_{\mathrm{SF}}(y) - 1 \geq l_{\mathrm{H}}(y).$$

Therefore, probabilities and Huffman-Kraft sums of wins, losses, and ties are bounded as

$$\begin{aligned}
&P(y) \geq 2 \cdot 2^{-l_{\mathrm{H}}(y)} = 2K(y) && \text{if } y \in W \\
&P(y) < 2^{-l_{\mathrm{H}}(y)} = K(y) && \text{if } y \in L \\
&K(y) = 2^{-l_{\mathrm{H}}(y)} \leq P(y) < 2 \cdot 2^{-l_{\mathrm{H}}(y)} = 2K(y) && \text{if } y \in T.
\end{aligned}$$

Thus, for any nonempty subset $A \subseteq S$,

$$\begin{aligned}
&P(A) \geq 2K(A) && \text{if } A \subseteq W && (13)\\
&P(A) < K(A) && \text{if } A \subseteq L && (14)\\
&K(A) \leq P(A) < 2K(A) && \text{if } A \subseteq T, && (15)
\end{aligned}$$

since $P(A) = \sum_{y \in A} P(y)$ and $K(A) = \sum_{y \in A} K(y)$ for any subset $A \subseteq S$.

Suppose that $L = \varnothing$. Then from (13) and (15) we have $P(y) \geq K(y)$ for all $y \in S$. Since $K(y)$ is an integer power of $1/2$ for all $y \in S$, there exists at least one element $y \in S$ with $P(y) > K(y)$, or else the source would be dyadic. But then

$$1 = \sum_{y \in S} P(y) > \sum_{y \in S} K(y) = 1,$$

which is a contradiction. Thus, in fact $L \neq \varnothing$, and therefore $P(L) > 0$. This implies $P(W) < 1$.

If $W = \varnothing$ then $P(W) - P(L) < 0$, and we are done. Suppose $W \neq \varnothing$. By Lemma 4.4, there exist Huffman-Kraft partitions $W_1, W_2, \ldots$ and $L_1, L_2, \ldots$ of W and L, respectively. Let k be the smallest integer such that $W_k \neq \varnothing$.

If $k = 1$, then $W_1 \neq \varnothing$, so $K(W_1) = \frac{1}{2}$ and therefore we get the contradiction that $1 > P(W_1) \geq 2K(W_1) = 1$ by (13). Thus $k \geq 2$.

We will use the following fact in the remainder of the proof. If $A \subseteq S$ and $K(A) = 2^{-(k-1)}$, then

$$\begin{aligned} P(A) &\geq P(W_k) && (16)\\ &\geq 2K(W_k) && (17)\\ &= 2^{-(k-1)} && (18)\\ &= K(A) && (19) \end{aligned}$$

where (16) follows from strong monotonicity; (17) follows from (13); and (18) follows from $P(W_k) = 2^{-k}$ since $W_k \neq \varnothing$. Also, $L_{k-1} = \varnothing$, for otherwise $K(L_{k-1}) = 2^{-(k-1)}$, which by (19) would imply $P(L_{k-1}) \geq K(L_{k-1})$, contradicting $P(L_{k-1}) < K(L_{k-1})$ by (14), as $L_{k-1} \subseteq L$.

Suppose $K(L) < 2^{-(k-2)}$. Then since $L_{k-1} = \varnothing$, at least the first $k-1$ bits in the binary expansion of $K(L)$ are zero, so $K(L) < 2^{-(k-1)}$. By Lemma 4.6, (where i, j, U, A, B, in Lemma 4.6 respectively correspond to 0, $k-2$, S, $W_k \cup L$, A here), since $0 < K(W_k \cup L) < 2^{-k} + 2^{-(k-1)} < 2^{-(k-2)} \leq 1 = K(S)$, there exists a subset $A \subseteq S - (W_k \cup L)$ such that $K(A) = 2^{-(k-2)} - K(W_k \cup L)$. Then

$$\begin{aligned} K(L \cup A) &= K(W_k \cup L \cup A) - K(W_k)\\ &= K(W_k \cup L) + K(A) - K(W_k)\\ &= 2^{-(k-2)} - K(W_k)\\ &= 2^{-(k-2)} - 2^{-k}\\ &= 2^{-(k-1)} + 2^{-k} \end{aligned}$$

so the Huffman-Kraft partition of $L \cup A$ provided by Lemma 4.4 consists of 2 disjoint subsets, E and F, of $L \cup A$ such that $E \cup F = L \cup A$, along with $K(E) = 2^{-k}$ and $K(F) = 2^{-(k-1)}$. Then $P(E) > 0$, and $P(F) \geq K(F)$ by (19).

Since $L \subseteq L \cup A = E \cup F$, we have $S - (W_k \cup E \cup F) \subseteq S - L \subseteq W \cup T$. Therefore, $P(S - (W_k \cup E \cup F)) \geq K(S - (W_k \cup E \cup F)) = 1 - 2^{-(k-2)}$ by (13) and (15). But we also have

$$\begin{aligned} P(S - (W_k \cup E \cup F)) &= 1 - P(W_k) - P(E) - P(F)\\ &< 1 - P(W_k) - P(F)\\ &\leq 1 - 2K(W_k) - K(F) && (20)\\ &= 1 - 2^{-(k-1)} - 2^{-(k-1)}\\ &= 1 - 2^{-(k-2)}\\ &= K(S - (W_k \cup E \cup F)), \end{aligned}$$

which is a contradiction, where (20) follows from (13). Therefore, our assumption was false that $K(L) < 2^{-(k-2)}$, so in fact $K(L) \geq 2^{-(k-2)}$. Thus, $k \neq 2$, for otherwise $K(L) \geq 1$ which contradicts $W \neq \varnothing$. Therefore, $k \geq 3$. Since $W_1 = \cdots = W_{k-1} = \varnothing$, we have $2K(W) < 2^{-(k-2)} \leq K(L)$, and so Lemma 4.7 shows $P(W) < P(L)$. ■

Theorem 5.1 guarantees that for each n, and for all non-dyadic sources of size n, Huffman codes always strictly competitively dominate Shannon-Fano codes. On the other hand, we saw in Section 3 that as the source size grows, an increasingly large fraction of non-dyadic sources have prefix codes that strictly competitively dominate Huffman codes. Said more casually, Huffman codes usually are dominated by another code but always dominate Shannon-Fano codes.

6. Bound on competitive advantage over Huffman codes

In this section, we derive an upper bound on the competitive advantage of an arbitrary prefix code over a Huffman code for a given source, and show that for every source of size at least four the upper bound can be approached arbitrarily closely by some sources. We show that no prefix code can have a competitive advantage of $\frac{1}{3}$ or higher over any Huffman code (Theorem 6.6), and in fact this upper bound is tight in that it can be approached arbitrarily closely from below for all source sizes, by at least some sources (Theorem 6.7).

If A is a subset of a source's alphabet, then at least one of the following three Huffman-Kraft sum conditions is satisfied: (i) A is empty and $K(A) = 0$; (ii) A is the entire alphabet and $K(A) = 1$; or (iii) $K(A)$ is a finite sum of negative integer powers of 2, and has a binary expansion of the form $K(A) = 0.b_1 b_2 \ldots$, with $b_i \in \{0, 1\}$ for all i, and where the number of nonzero bits is at least one and is finite.

If A is a subset of a source alphabet, then the probability $P(A)$ and the Huffman-Kraft sum $K(A)$ are related. When A is empty, $P(A) = K(A) = 0$, and when A is the entire source alphabet, $P(A) = K(A) = 1$. If the source is dyadic, then $P(A) = K(A)$ is always true. For non-dyadic sources, the relationship between $P(A)$ and $K(A)$ is more complicated. We next establish some lemmas that are used to prove Theorem 6.6. Some of these lemmas relate the probabilities and the Huffman-Kraft sums of source alphabet subsets.

In this section, for any given source, if C is a prefix code that competes against a Huffman code for the same source, then define the events W (i.e., C "wins"), L (i.e., C "loses"), and T (i.e., C "ties"), as in (1) (taking $C_1 = C$, and C_2 as the Huffman code).

In what follows, the definitions of W, L, and T from (1) (taking C_2 as the Huffman code) are used in Lemma 6.1, Lemma 6.2, and Theorem 6.6.

Lemma 6.1. *For any source, if a prefix code C has a positive competitive advantage over a Huffman code, then for at least one source symbol, C produces a longer codeword than the Huffman codeword.*

Proof. It suffices to show that L is nonempty. Since C has a positive competitive advantage over a Huffman code H, we have $P(W) > P(L)$. Then,

$$
\begin{aligned}
0 &\leq E[l_C(X)] - E[l_H(X)] && (21)\\
&= \sum_{y \in W} (l_C(y) - l_H(y))\, P(y) + \sum_{y \in L} (l_C(y) - l_H(y))\, P(y) && (22)\\
&< \sum_{y \in L} (l_C(y) - l_H(y))\, P(y) && (23)
\end{aligned}
$$

where (21) follows since a prefix code C cannot have a lower expected length than the Huffman code for a given source; (22) follows since $l_C(y) - l_H(y) = 0$ for all $y \in T$; and (23) follows since $l_C(y) - l_H(y) < 0$ for all $y \in W$ and $P(W) > P(L) \geq 0$, so $W \neq \varnothing$. If $L = \varnothing$, then (23) would yield a contradiction. ∎

Lemma 6.2. *For any source, if a prefix code C has a positive competitive advantage over a Huffman code, then the Huffman-Kraft sums of the set W of wins and set L of losses of C satisfy $K(W) < K(L)$.*

Proof. Let H denote the Huffman code and suppose, to the contrary, that

$$K(W) \geq K(L). \tag{24}$$

Let S be the source's alphabet and let $T = S - (W \cup L)$ be the set of ties. Then

$$\begin{aligned}
1 &\geq \sum_{x\in S} 2^{-l_C(x)} && (25)\\
&= \sum_{x\in W} 2^{-l_C(x)} + \sum_{x\in T} 2^{-l_C(x)} + \sum_{x\in L} 2^{-l_C(x)}\\
&> \sum_{x\in W} 2^{-l_C(x)} + \sum_{x\in T} 2^{-l_C(x)} && (26)\\
&= \sum_{x\in W} 2^{-l_C(x)} + \sum_{x\in T} 2^{-l_H(x)} && (27)\\
&\geq \sum_{x\in W} 2^{-l_H(x)+1} + \sum_{x\in T} 2^{-l_H(x)} && (28)\\
&= 2K(W) + K(T)\\
&\geq K(W) + K(T) + K(L) && (29)\\
&= 1, && (30)
\end{aligned}$$

a contradiction, where (25) is the Kraft inequality (1.1) applied to the prefix code C; (26) follows from $L \neq \varnothing$ by Lemma 6.1; (27) follows from $l_C(x) = l_H(x)$ when $x \in T$; (28) follows from $l_C(x) \leq l_H(x) - 1$ when $x \in W$; (29) follows from (24); and (30) follows since the Huffman tree is complete. ■

Lemma 6.3. *If $b > a$, then $\frac{x+a}{x+b}$ is monotonically increasing in x for all $x \neq -b$.*

Proof. $\frac{d}{dx}\left(\frac{x+a}{x+b}\right) = \frac{b-a}{(x+b)^2} > 0$. ■

Lemma 6.4. *Given a source with alphabet S, let U and V be disjoint subsets of S with Huffman-Kraft sums satisfying $K(U) < 2^{-i} \leq K(V)$ for some integer $i \geq 0$. Then $P(U) < 2P(V)$.*

Proof. Since $K(V) \geq 2^{-i}$, we have $P(V) > 0$. If $K(U) = 0$ then $P(U) = 0 < 2P(V)$, so we may assume $K(U) > 0$.

By Lemma 4.4, there exists a Huffman-Kraft partition $V_1, V_2, \ldots$ of V. Let $k \geq 0$ be the smallest integer such that $V_k \neq \varnothing$. Since $K(V) \geq 2^{-i}$, we have $k \leq i$. By Lemma 4.6 (taking $B = U'$, $U = S$, $A = U$, and $j = k$), there exists a subset $U' \subseteq S - U$ such that $K(U') = 2^{-k} - K(U) > 0$. Then $|U \cup U'| \geq 2$ since both U and U' are nonempty, and also $K(U \cup U') = 2^{-k} = K(V_k)$, so $P(U \cup U') \leq 2P(V_k)$ by Lemma 4.5 (taking $A = U \cup U'$, $B = V_k$, and $i = k$). Therefore

$$P(U) \leq 2P(V_k) - P(U') < 2P(V_k) \leq 2P(V).$$

■

For any Huffman-Kraft partition $A_1, A_2, \ldots$ of a subset A of a source alphabet and for any integer $k \geq 1$ we will define the notation $A_{<k} = A_1 \cup \cdots \cup A_{k-1}$ and $A_{\leq k} = A_1 \cup \cdots \cup A_k$, as well as $A_{>k} = A_{k+1} \cup A_{k+2} \ldots$ and $A_{\geq k} = A_k \cup A_{k+1} \ldots$.

Lemma 6.5. *Given a source with alphabet S, suppose $U, V \subseteq S$ are disjoint subsets with Huffman-Kraft sums satisfying $K(U) < K(V)$. Then $P(U) - P(V) < \frac{1}{3}$.*

Proof. If $K(U) = 0$ then $P(U) - P(V) \leq 0 < \frac{1}{3}$, so suppose $K(U) > 0$. By Lemma 4.4, there exist Huffman-Kraft partitions $U_1, U_2, \dots$ of U and $V_1, V_2, \dots$ of V. Let $m \geq 0$ be the smallest integer such that $V_m \neq \varnothing$ and $U_m = \varnothing$; such an integer exists because $K(U) < K(V)$. Since $P(U) - P(V) \leq P(U) - P(V_{\leq m})$, and U and $V_{\leq m}$ are disjoint, without loss of generality we will assume $V_i = \varnothing$ for all $i > m$.

We now assert that for certain nonnegative integers $k \leq m - 1$, there exists $A \subseteq S$ satisfying the following three conditions:

$$K(A) = 2^{-k} \tag{31}$$

$$U_{\geq k+1} \cup V_{\geq k+1} \subseteq A \tag{32}$$

$$P(U_{\geq k+1}) - P(V_{\geq k+1}) < \frac{1}{3} P(A). \tag{33}$$

Specifically, we will first show that this assertion is true when $k = m - 1$. Then, we will show inductively that whenever the assertion is true for some positive k it must also be true for some smaller nonnegative k. We then will conclude that the assertion must be true for $k = 0$.

Once we have established the assertion is true for $k = 0$, we can further infer that

$$P(U) - P(V) = P(U_{\geq 1}) - P(V_{\geq 1}) < \frac{1}{3} P(A) = \frac{1}{3},$$

where $P(A) = 1$ since $K(A) = 1$, thus proving the lemma.

Base Step: $k = m - 1$. Since $K(S - (U_{<m} \cup V_{<m})) = 1 - K(U_{<m} \cup V_{<m})$ is an integer multiple of $2^{-(m-1)}$, and

$$0 < K(V_m) \leq K(U_{\geq m} \cup V_{\geq m}) = K(U_{>m}) + K(V_m) < 2^{-(m-1)},$$

Lemma 4.6 shows (taking $A \leftarrow U_{\geq m} \cup V_{\geq m}$, $B \leftarrow U'$, $U \leftarrow S - (U_{<m} \cup V_{<m})$, $i = j \leftarrow m - 1$) there exists $U' \in= S - (U \cup V)$ such that $K(U' \cup U_{\geq m} \cup V_{\geq m}) = 2^{-(m-1)}$. Then setting

$$A = U' \cup U_{\geq m} \cup V_{\geq m} \tag{34}$$

gives

$$K(A) = 2^{-(m-1)}. \tag{35}$$

Also, since $K(U_{\geq m}) = K(U_{>m}) < 2^{-m} = K(V_m) = K(V_{\geq m})$, Lemma 6.4 shows $P(U_{\geq m}) < 2P(V_{\geq m})$. Therefore,

$$\begin{aligned} \frac{P(U_{\geq m}) - P(V_{\geq m})}{P(A)} &< \frac{P(U_{\geq m}) - P(V_{\geq m})}{P(U_{\geq m}) + P(V_{\geq m})} && (36)\\ &< \frac{P(V_{\geq m})}{3P(V_{\geq m})} && (37)\\ &= \frac{1}{3}, && (38) \end{aligned}$$

where (36) follows from $P(U') > 0$ since $K(U') = K(A) - K(U_{\geq m} \cup V_{\geq m}) > 2^{-(m-1)} - 2^{-(m-1)} = 0$;

and (37) follows from $P(U_{\geq m}) < 2P(V_{\geq m})$ and Lemma 6.3.

By (35), (34), and (38), these conditions hold for $k = m-1$.

Inductive Step: $1 \leq k \leq m-1$. Notice that if $m = 1$, then the base case of $k = m-1$ automatically proves the assertion for $k = 0$, so we may assume $m \geq 2$.

Assume that (31) – (33) hold for some positive integer $k \leq m-1$ and for some $A \subseteq S$. We will show that there exists $A' \subseteq S$ that satisfies the three conditions above for some index $j \in \{0, \ldots, k-1\}$.

From the definition of m, we have $K(U_i) = K(V_i)$ for all $i < m$. In particular, $K(U_k) = K(V_k)$ since $k \leq m-1$. Then $K(U_k) + K(V_k) \in \{0, 2^{-(k-1)}\}$, so $K(U_{\leq k} \cup V_{\leq k}) = K(U_{<k} \cup V_{<k}) + K(U_k \cup V_k)$ is an integer multiple of $2^{-(k-1)}$. Therefore, $K(S - U_{\leq k} \cup V_{\leq k}) = 1 - K(U_{\leq k} \cup V_{\leq k})$ is an integer multiple of $2^{-(k-1)}$. Since $0 < K(A) = 2^{-k} < 2^{-(k-1)}$, and $A \subseteq S - U_{\leq k} \cup V_{\leq k}$, Lemma 4.6 shows (taking $A \leftarrow A$, $B \leftarrow B$, $U \leftarrow S - (U_{\leq m} \cup V_{\leq m})$, $i = j \leftarrow k-1$) there exists $B \subseteq S - (A \cup U_{\leq k} \cup V_{\leq k})$ with $K(B) = 2^{-(k-1)} - K(A) = 2^{-k}$.

Case 1: $K(U_k) = K(V_k) = 0$. Let $A' = A \cup B$. Then $K(A') = K(A) + K(B) = 2^{-(k-1)}$ and

$$U_{\geq k} \cup V_{\geq k} = U_{\geq k+1} \cup V_{\geq k+1} \subseteq A \subseteq A'$$

from (32). Also,

$$P(U_{\geq k}) - P(V_{\geq k}) = P(U_{\geq k+1}) - P(V_{\geq k+1}) < \frac{1}{3}P(A) < \frac{1}{3}P(A')$$

from (33) and $K(A) = K(A') - K(B) < K(A')$. Thus A' satisfies conditions (31) – (33) except the index k has been reduced to $j = k-1 \geq 0$.

Case 2: $K(U_k) = K(V_k) = 2^{-k}$. Let $j \leq k$ be the smallest integer such that $K(U_i) = 2^{-i} = K(V_i)$ for all $i \in \{j, \ldots, k\}$. Then $j \geq 2$, since otherwise, $j = 1$ would imply $1 \geq K(U) + K(V) > 2K(U) \geq 2K(U_1) = 1$, a contradiction. Also, from the definition of j, we have $K(U_{j-1}) = K(V_{j-1}) = 0$. Let

$$A' = A \cup B \cup (U_{\geq j} - U_{>k}) \cup (V_{\geq j} - V_{>k}). \tag{39}$$

Then

$$\begin{aligned}
U_{\geq j-1} \cup V_{\geq j-1} &= U_{\geq j} \cup V_{\geq j} \\
&= (U_{\geq j} - U_{>k}) \cup (V_{\geq j} - V_{>k}) \cup U_{\geq k+1} \cup V_{\geq k+1} \\
&\subseteq (U_{\geq j} - U_{>k}) \cup (V_{\geq j} - V_{>k}) \cup A \qquad (40) \\
&\subseteq A' \qquad (41)
\end{aligned}$$

where (40) follows from (32); and

$$\begin{aligned}
K(A') &= K(A) + K(B) + K(U_{\geq j} - U_{>k}) + K(V_{\geq j} - V_{>j}) \\
&= 2^{-k} + 2^{-k} + 2\sum_{i=j}^{k} 2^{-i} \\
&= 2^{-(j-2)}. \qquad (42)
\end{aligned}$$

Now let $E \in \{A, B\}$ such that $P(E) = \min(P(A), P(B))$. Note $K(E) = 2^{-k}$, since $K(A) =$

$K(B) = 2^{-k}$. Since

$$K((V_{\geq j} - V_{>k}) \cup E) = \sum_{i=j}^{k} 2^{-i} + K(E) = 2^{-(j-1)} - 2^{-k} + 2^{-k} = 2^{-(j-1)} > 2^{-j} = K(U_j),$$

strong monotonicity of Huffman-Kraft sums and Lemma 4.3 imply $P(U_j) \leq P(V_{\geq j} - V_{>k}) + P(E)$.

Suppose $j < k$. Then for all $i \in \{j+1, \ldots, k\}$, we have $P(U_i) \leq P(V_{i-1})$ by strong monotonicity of Huffman-Kraft sums, since $K(U_i) = 2^{-i} < 2^{-(i-1)} = K(V_{i-1})$. Thus

$$\begin{aligned} P(U_{\geq j+1} - U_{>k}) &= P(U_{j+1}) + \cdots + P(U_k) \\ &\leq P(V_j) + \cdots + P(V_{k-1}) \\ &= P(V_{\geq j} - V_{>k-1}). \end{aligned}$$

Therefore

$$\begin{aligned} P(U_{\geq j} - U_{>k}) &= P(U_j) + P(U_{\geq j+1} - U_{>k}) \\ &\leq P(V_{\geq j} - V_{>k}) + P(E) + P(V_{\geq j} - V_{>k-1}) \\ &= 2P(V_{\geq j} - V_{>k-1}) + P(V_k) + P(E). \end{aligned} \tag{43}$$

On the other hand, suppose $j = k$. Then $V_{\geq j} - V_{>k-1} = \varnothing$, so we also have

$$\begin{aligned} P(U_{\geq j} - U_{>k}) &= P(U_j) \\ &\leq P(V_{\geq j} - V_{>k}) + P(E) \\ &= 2P(V_{\geq j} - V_{>k-1}) + P(V_k) + P(E). \end{aligned} \tag{44}$$

Now we combine both the cases $j < k$ and $j = k$. Since $V_m \subseteq A$ and $K(V_m) = 2^{-m} < 2^{-k} = K(A)$, neither V_m nor $A - V_m$ is empty, so $|A| = |V_m| + |A - V_m| \geq 2$. Then since $K(A) = 2^{-k} = K(V_k)$, Lemma 4.5 shows $P(A) \leq 2P(V_k)$. Thus

$$P(E) = \min(P(A), P(B)) \leq \min(2P(V_k), P(B)),$$

and so

$$2P(V_k) + P(E) + P(B) \geq 3P(E). \tag{45}$$

Finally, we apply Lemma 6.3 by using the values

$$\begin{aligned} x &= P(U_{\geq j} - U_{>k}) \\ a &= -P(V_{\geq j} - V_{>k}) \\ b &= P(V_{\geq j} - V_{>k}) + P(B) \\ x' &= 2P(V_{\geq j} - V_{>k-1}) + P(V_k) + P(E). \end{aligned}$$

It is clear that $a < b$, and we have $x \le x'$ from (43) (for $j < k$) and (44) (for $j = k$), Thus,

$$\begin{aligned}
\frac{x+a}{x+b} &\le \frac{x'+a}{x'+b} \\
\frac{P(U_{\ge j} - U_{>k}) - P(V_{\ge j} - V_{>k})}{P(U_{\ge j} - U_{>k}) + P(V_{\ge j} - V_{>k}) + P(B)} &\le \frac{P(V_{\ge j} - V_{>k-1}) + P(E)}{3P(V_{\ge j} - V_{>k-1}) + 2P(V_k) + P(E) + P(B)} \\
&\le \frac{P(V_{\ge j} - V_{>k-1}) + P(E)}{3P(V_{\ge j} - V_{>k-1}) + 3P(E)} \qquad (46) \\
&= \frac{1}{3}, \qquad (47)
\end{aligned}$$

where (46) follows from (45). Therefore,

$$\begin{aligned}
P(U_{\ge j-1}) - P(V_{\ge j-1}) &= P(U_{\ge j}) - P(V_{\ge j}) \\
&= P(U_{\ge j} - U_{>k}) - P(V_{\ge j} - V_{>k}) + P(U_{\ge k+1}) - P(V_{\ge k+1}) \\
&< \frac{1}{3}(P(U_{\ge j} - U_{>k}) + P(V_{\ge j} - V_{>k}) + P(B) + P(A)) \qquad (48) \\
&= \frac{1}{3}P(A'), \qquad (49)
\end{aligned}$$

where (48) follows from (47) and (33); and (49) follows from (39). Thus A' satisfies the three conditions (31) – (32) by way of (42), (41), and (49), except the index k has been reduced to $j-2 \in \{0, \dots, k-2\}$. ∎

Theorem 6.6. *For any source, the competitive advantage of any prefix code over a Huffman code is less than* $\frac{1}{3}$.

Proof. Let C denote an arbitrary prefix code for the source. Let W and L denote the sets of wins and losses, respectively, of C over the Huffman code. It suffices to assume the competitive advantage of C over the Huffman code is positive, so $W \ne \varnothing$. Then Lemma 6.1 implies $L \ne \varnothing$, and Lemma 6.2 implies $K(W) < K(L)$. Therefore, $P(W) - P(L) < \frac{1}{3}$ by Lemma 6.5. ∎

The following theorem shows that for any size at least four, sources can be found whose competitive advantages over Huffman codes are arbitrarily close to $1/3$ and whose average lengths are arbitrarily close to that of a Huffman code.

Theorem 6.7. *For every $n \ge 4$, there exists a source of size n and a prefix code that has a competitive advantage over a Huffman code arbitrarily close to $\frac{1}{3}$ and the code's average length is arbitrarily close to that of the Huffman code.*

Proof. Let $n \ge 4$ and $\epsilon > 0$, and define $\alpha = \frac{\epsilon/2}{1-2^{-n+3}}$. Let the source be of size n and with symbol probabilities:

$$\begin{aligned}
p_1 &= \frac{1}{3} + \epsilon \\
p_2 &= \frac{1}{3} \\
p_3 &= \frac{1}{3} - 2\epsilon \\
p_k &= \alpha 2^{4-k} \qquad (4 \le k \le n).
\end{aligned}$$

One can verify that $p_1 + \cdots + p_n = 1$ and for each $k \in \{2, \ldots, n\}$ we have $p_k > p_{k+1} + \cdots + p_n$, so the Huffman code for the source assigns a word of length k to p_k for $k = 1, \ldots, n-1$, and also a word of length $n-1$ to p_n.

Define a prefix code C which is identical to the Huffman code, except that it reassigns p_1, p_2, and p_3 to codewords of lengths 3, 1, and 2, respectively. The code C will produce a shorter codeword than that of the Huffman code with probability $p_2 + p_3$ and will produce a longer codeword with probability p_1. Thus, the competitive advantage of C over the Huffman code is $\Delta = p_2 + p_3 - p_1 = \frac{1}{3} - 3\epsilon$.

Denote the codeword lengths of the Huffman code by l_i. The average length of the Huffman code is

$$1 \cdot \left(\frac{1}{3} + \epsilon\right) + 2 \cdot \left(\frac{1}{3}\right) + 3 \cdot \left(\frac{1}{3} - 2\epsilon\right) + \sum_{k=4}^{n} p_i l_i$$

and the average length of C is

$$1 \cdot \left(\frac{1}{3}\right) + 2 \cdot \left(\frac{1}{3} - 2\epsilon\right) + 3 \cdot \left(\frac{1}{3} + \epsilon\right) + \sum_{k=4}^{n} p_i l_i$$

so their difference is $-\frac{14}{3}\epsilon$.

In summary, the code C achieves a competitive advantage over the Huffman code of $\frac{1}{3} - 3\epsilon$ and has an average length at most $\frac{14}{3}\epsilon$ greater than that of the Huffman code. Taking ϵ arbitrarily small makes the competitive advantage approach $\frac{1}{3}$ and the average length difference approach zero. ■

7. Bound on competitive advantage over Shannon-Fano codes

On one hand, Shannon-Fano codes are efficient, since they suffice in proving Shannon's source coding theorem that says the average length of optimal block codes arbitrarily approaches from above the entropy of a source, as the block size grows. The proof uses the fact that the average length of a Shannon-Fano code is always within one bit of the source entropy, and so the average length per symbol of a Shannon-Fano code for a source block of size n is within $\frac{1}{n}$ bit of the source entropy.

One the other hand, Huffman codes are strictly better than Shannon-Fano codes in an average length sense for non-dyadic sources, and perform equally well for for dyadic sources. Similarly, in a competitive sense, Theorem 5.1 showed that Huffman codes strictly competitively dominate Shannon-Fano codes if and only if the source is not dyadic.

The competitive advantage of one code over a Shannon-Fano code (or, actually, any other code) is trivially upper bounded by one, and the average length of a code can be at most one bit less than that of a Shannon-Fano code. The following theorem shows that there exist increasingly large sources with prefix codes that can approach both of these extremes over Shannon-Fano codes simultaneously.

Theorem 7.1. *For every positive integer n, there exists a source of size n and a prefix code that has a competitive advantage of at least $1 - 2^{-n+2}$ over a Shannon-Fano code for the source, and the code's average length is at least $1 - 2^{-n+2}$ less than the average length of the Shannon-Fano code.*

Proof. Let $\epsilon \in (0, 4^{-n})$ and let X be a source of size n whose probabilities are

$$p_k = \begin{cases} 2^{-k} - \epsilon & \text{if } 1 \le k \le n-1 \\ 2^{-n+1} + (n-1)\epsilon & \text{if } k = n. \end{cases}$$

Since $\left\lceil \log_2 \frac{1}{p} \right\rceil = m$ if and only if $2^{-m} \le p < 2^{-m+1}$, a Shannon-Fano code for this distribution has codeword lengths

$$l_k = \left\lceil \log_2 \frac{1}{p_k} \right\rceil = \begin{cases} k+1 & \text{if } 1 \le k \le n-1 \\ n-1 & \text{if } k = n. \end{cases}$$

Note that l_n was determined from the fact that for all $n \ge 1$,

$$(n-1)4^{-n} < 2^{-n+1}. \tag{50}$$

Let C be a prefix code that assigns the word $1^{k-1}0$ to the outcomes that have probability p_k when $k < n$, and assigns the word 1^{n-1} to the outcome with probability p_n. This prefix code produces a shorter codeword than a Shannon-Fano code whenever $1 \le k \le n-1$, and ties when $k = n$, so its competitive advantage over a Shannon-Fano code is lower bounded as

$$\Delta = 1 - p_n = 1 - \frac{1}{2^{n-1}} - (n-1)\epsilon \ge 1 - \frac{1}{2^{n-1}} - \frac{n-1}{4^n} > 1 - \frac{1}{2^{n-2}}. \tag{51}$$

where (51) follows from (50).

The difference between the average lengths of the Shannon-Fano code and the code C is

$$(n-1)p_n + \sum_{k=1}^{n-1}(k+1)p_k - (n-1)p_n - \sum_{k=1}^{n-1} kp_k = \sum_{k=1}^{n-1} p_k = 1 - p_n, \tag{52}$$

the same quantity as the competitive advantage previously computed in (51).

■

In the preceding proof, the average length of the code C is at most 2^{-n+2} more than the source entropy, since

$$\begin{aligned} E[l_C(X))] &< E[l_{C_{SF}}(X))] - 1 + \frac{1}{2^{n-2}} && (53) \\ &< H(X) + 1 - 1 + \frac{1}{2^{n-2}} && (54) \\ &= H(X) + \frac{1}{2^{n-2}}. \end{aligned}$$

where (53) follows from (51) and (52); and (54) follows from Shannon's source coding theorem. We also note that the term 2^{-n+2} that occurs in the bounds of the theorem can be sharpened to be arbitrarily close to 2^{-n+1} but we chose to keep the proof simple instead.

8. Small codes

In this section, we analyze which sources of size at most 4 have competitively optimal Huffman codes.

Theorem 8.1. *Huffman codes are competitively optimal for all sources of size at most* 3*.*

Proof. If the source is of size 1 or 2, the result is trivial, so suppose the size is 3. Denote the source symbols by $1, 2, 3$ such that $P(1) \geq P(2) \geq P(3) > 0$. The word lengths of a Huffman code H are $l_H(1) = 1$ and $l_H(2) = l_H(3) = 2$. Let C denote any other prefix code, and use the notation W and L, as in (1). It is not possible for $1 \in W$, since $l_C(1) \geq 1 = l_H(1)$. If $2 \in W$, then $l_C(2) = 1$ and therefore $l_C(1), l_C(3) \geq 2$, so $1 \in L$ and $3 \notin W$, which implies the competitive advantage of C over the Huffman code is $\Delta = P(W) - P(L) = P(2) - P(L) \leq P(2) - P(1) \leq 0$. Alternatively, if $3 \in W$, then we similarly conclude $\Delta \leq 0$. Finally, if $2, 3 \notin W$, then $P(W) = 0$, so $\Delta \leq 0$. ■

Theorem 8.2. *Let Q be the hexahedron with vertices $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{2}{5}, \frac{1}{5}, \frac{1}{5})$, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{6})$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. For every source of size 4 with probabilities $p_1 \geq p_2 \geq p_3 \geq p_4 > 0$, a Huffman code is competitively optimal if the triple (p_1, p_2, p_3) lies in the exterior of Q, and is not competitively optimal if the triple lies in the interior of Q.*

Proof. Denote the source symbols by $1, 2, 3, 4$ and their probabilities by p_1, p_2, p_3, p_4, respectively. We will determine conditions on $p_1, \ldots, p_4$ such that there exists a prefix code with a positive competitive advantage over a Huffman code. It suffices to consider complete prefix codes, since any non-complete prefix code contains at least one codeword that could be shortened without decreasing its competitive advantage. The only possible codeword length distributions for such size-4 codes are $1, 2, 3, 3$ and $2, 2, 2, 2$. In either case, the Huffman algorithm merges the source symbols 3 and 4 to form a new symbol with probability $p_3 + p_4$.

Suppose $p_3 + p_4 > p_1$. Then the Huffman algorithm merges 1 and 2 and then the $(3, 4)$ symbol is merged with the $(1, 2)$ symbol to get a balanced tree with codeword lengths $2, 2, 2, 2$. If a size-4 prefix code achieves a positive competitive advantage over this Huffman code, then it must have codeword lengths $1, 2, 3, 3$, for otherwise only ties would occur. In this case, the competitive advantage would be the probability of the new code's length-1 word minus the sum of the probabilities of its two length-3 words, which equals $p_1 - (p_3 + p_4) < 0$, so in fact the new code would be strictly competitively dominated by the Huffman code. The competitive advantage would still not be positive even if $p_3 + p_4 = p_1$ and the Huffman algorithm created codewords with lengths $2, 2, 2, 2$.

Alternatively, assume $p_3 + p_4 \leq p_1$ with the Huffman algorithm merging the $(3, 4)$ symbol with 2, and then merging the resulting $(2, (3, 4))$ symbol with 1. The resulting codeword lengths are $1, 2, 3, 3$. The competitive advantage of any depth-2 balanced tree over the Huffman code would be $p_3 + p_4 - p_1 \leq 0$, so such codes are competitively dominated by the Huffman code. Thus, any code C with a positive competitive advantage Δ over the Huffman code must have lengths $1, 2, 3, 3$, and hence C just permutes the Huffman code's assignment of codeword lengths to source symbols.

Suppose $l_C(1) = 1$. If $l_C(2) = 2$, then $\Delta = 0$. If $l_C(2) = 3$ and $l_C(3) = 2$, then $\Delta = p_3 - p_2 \leq 0$. If $l_C(2) = 3$ and $l_C(4) = 2$, then $\Delta = p_4 - p_2 \leq 0$.

Alternatively, suppose $l_C(1) \neq 1$. There are 9 possible cases for $(l_C(1), l_C(2), l_C(3), l_C(4))$:
(2,1,3,3): $\Delta = p_2 - p_1 \leq 0$
(2,3,1,3): $\Delta = p_3 - p_1 - p_2 \leq 0$
(2,3,3,1): $\Delta = p_4 - p_1 - p_2 \leq 0$
(3,2,1,3): $\Delta = p_3 - p_1 \leq 0$
(3,2,3,1): $\Delta = p_4 - p_1 \leq 0$
(3,3,1,2): $\Delta = p_3 + p_4 - p_1 - p_2 \leq 0$

(3,3,2,1): $\Delta = p_3 + p_4 - p_1 - p_2 \le 0$
(3,1,2,3): $\Delta = p_2 + p_3 - p_1$
(3,1,3,2): $\Delta = p_2 + p_4 - p_1$.
So the only codes C that can yield $\Delta > 0$ are the cases $(3,1,2,3)$ and $(3,1,3,2)$.

Let us denote the following inequalities:

$$\begin{aligned}
(I1):&\ p_1 \ge p_2\\
(I2):&\ p_2 \ge p_3\\
(I3):&\ p_3 \ge p_4\\
(I4):&\ p_4 > 0\\
(I5):&\ p_3 + p_4 \le p_1\\
(I6):&\ p_2 + p_3 > p_1\\
(I7):&\ p_2 + p_4 > p_1.
\end{aligned}$$

Inequalities (I1) – (I6) determine a set in $\mathbb{R}^3$ whose interior is a hexahedron specified by the 5 vertices $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{2}{5}, \frac{1}{5}, \frac{1}{5})$, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{6})$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. The first 3 vertices satisfy $\sim$(I7) with equality, the 4^{th} vertex satisfies (I7), and 5^{th} vertex satisfies $\sim$(I7). Therefore, the hexahedron is cut into two tetrahedra by (I7) and is known as a triangular dipyramid.

The Huffman code is competitively optimal in the exterior of this hexahedron, is not competitively optimal in the interior of this hexahedron, and is sometimes competitively optimal on the boundary. ■

Corollary 8.3. *If a source of size* 4 *is chosen uniformly at random from a flat Dirichlet distribution, then the probability its Huffman code is competitively optimal is* $2/3$*.*

Proof. The hexahedron in Theorem 8.2 is a union of two tetrahedra, whose volumes are computed using determinants as (e.g., [13])

$$\frac{1}{6}\cdot\begin{vmatrix}1/2 & 1/2 & 0 & 1\\ 2/5 & 1/5 & 1/5 & 1\\ 1/3 & 1/3 & 1/3 & 1\\ 1/3 & 1/3 & 1/6 & 1\end{vmatrix} = \frac{1}{6}\cdot\frac{1}{180} \qquad\qquad \frac{1}{6}\cdot\begin{vmatrix}1/2 & 1/2 & 0 & 1\\ 2/5 & 1/5 & 1/5 & 1\\ 1/3 & 1/3 & 1/3 & 1\\ 2/5 & 1/5 & 1/5 & 1\end{vmatrix} = \frac{1}{6}\cdot\frac{1}{120}\ .$$

The set of all p_1, p_2, p_3, p_4 satisfying $p_1 \ge p_2 \ge p_3 \ge p_4 > 0$ and $p_1 + p_2 + p_3 + p_4 = 1$ is determined by the 4 inequalities

$$\begin{aligned}
p_1 &\ge p_2\\
p_2 &\ge p_3\\
p_1 + p_2 + 2p_3 &\ge 1\\
p_1 + p_2 + p_3 &< 1.
\end{aligned}$$

These form a tetrahedron with vertices $(1,0,0)$, $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ whose volume is

$$\frac{1}{6}\cdot\begin{vmatrix}1 & 0 & 0 & 1\\ 1/2 & 1/2 & 0 & 1\\ 1/3 & 1/3 & 1/3 & 1\\ 1/4 & 1/4 & 1/4 & 1\end{vmatrix} = \frac{1}{6}\cdot\frac{1}{24}\ .$$

Thus the probability of randomly selecting a source from a flat Dirichlet distribution whose Huffman code is not competitively optimal is $(\frac{1}{180} + \frac{1}{120})/\frac{1}{24} = \frac{1}{3}$. So the probability the Huffman code is competitively optimal is $\frac{2}{3}$. ■

9. Experimental evidence

We demonstrate numerically that if a source is chosen at random, then as the source size grows, the probability becomes nearly zero that a Huffman code will be competitively optimal. Experimentally, this probability is less than 1% when the source size is at least 20. That is, with near certainty each Huffman code will be competitively dominated by some other prefix code as the source size increases. This indicates that for most sources, from a competitive advantage point of view, there really is no "best" code to use. Each code can be strictly competitively dominated by another in never-ending cycles of code sequences.

One way to generate source probabilities $p_1, \ldots, p_n$ chosen according to a flat Dirichlet distribution is to choose n points independently and uniformly on a circle of circumference 1 and then use the n distances between neighboring points as the desired probabilities. Such a procedure treats all sources equally and indeed yields interesting results.

For any source of size n, exhaustively checking whether each complete prefix code competitively dominates the Huffman code appears to become a computationally infeasible task as n grows, since the number of such prefix codes grows quickly. However, Lemma 3.4 gives a sufficient condition for a Huffman tree to not be competitively optimal, which allows us to obtain a lower bound on the probability that a Huffman code is not competitively optimal for a given source. Thus we can randomly select many sources and determine if such a condition holds, in which case we can then declare the Huffman code not competitively optimal. This suboptimal condition turns out to be overwhelmingly sufficient to observe that the probability is practically zero that the Huffman code of a randomly chosen source is competitively optimal even for relatively small source sizes.

For each source size $n \in \{3, \ldots, 34\}$, we generated 10^6 sources from a flat Dirichlet distribution, i.e., chosen uniformly at random on the $(n-1)$-dimensional simplex embedded in $\mathbb{R}^n$. For each such source we determined whether the sufficient condition of Lemma 3.4 was satisfied. Fig 2 plots for each n the fraction of the randomly generated sources that satisfied the sufficient condition. That is, the true fraction of the randomly generated sources for which a Huffman code was not competitively optimal lies above the plotted curve. The observed lower bound curve quickly tends toward 1, so the true fraction of the randomly generated sources with competitively non-optimal Huffman codes tends toward 1 as well.

For the case of $n = 3$, Theorem 8.1 guarantees that 100% of the randomly chosen sources will have competitively optimal Huffman codes, which is exactly what was observed experimentally.

For the case $n = 4$, Corollary 8.3 gives a $2/3$ probability of a randomly chosen source to have a competitively optimal Huffman code. The experimentally observed upper bound was 66.6992%.

For $n \geq 5$, one can see that the probability a randomly chosen source has a competitively optimal Huffman code rapidly decreases towards 0, and in fact no such competitively optimal Huffman codes were observed out of the million chosen for each $n \geq 31$.

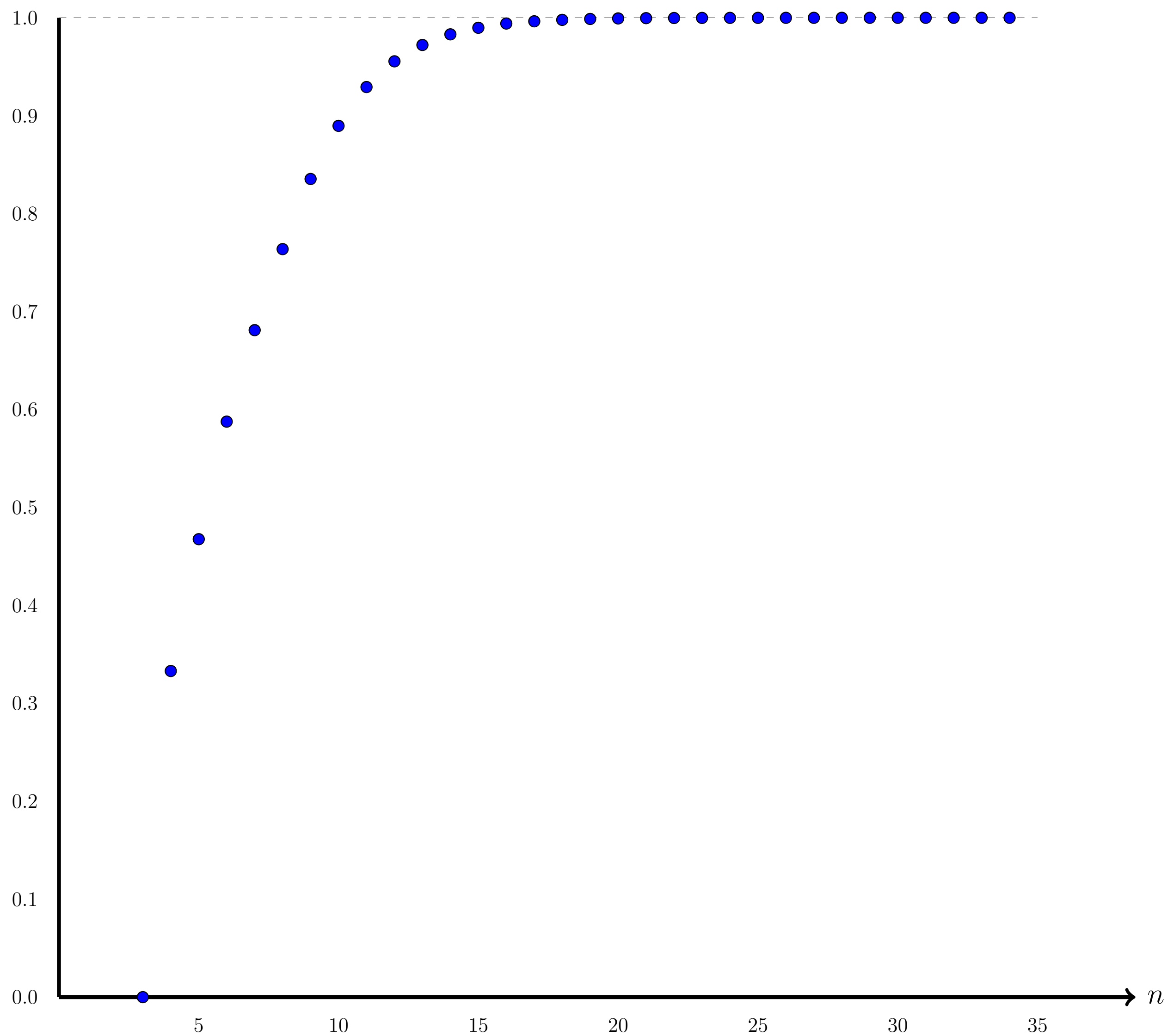

Figure 2: Lower bound on the fraction of 10^6 randomly chosen sources whose Huffman code is not competitively optimal, as a function of the source size n. For $n = 15$ Huffman codewords, about 99% of randomly selected sources did not have competitively optimal Huffman codes. For $n \geq 31$, all 10^6 randomly chosen sources had Huffman codes that were not competitively optimal.

Acknowledgment: The authors thank UCSD undergraduate student Marco Bazzani for some helpful discussions.

References

[1] R. Bell and T. M. Cover, "Competitive optimality of logarithmic investment", *Mathematical Operations Research*, vol. 5, no. 2, pp. 161 – 166, May 1980.

[2] J. Berstel, D. Perrin, and C. Reutenauer, *Codes and Automata*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2009.

[3] J.R. Bhatnagar, "Competitive optimality: A novel application in evaluating practical AI Systems", *Engineering Applications of Artificial Intelligence*, vol. 102, article 104241, June 2021.

[4] P. Billingsley, *Probability and Measure*, John Wiley & Sons, New York, 1986.

[5] S. Congero and K. Zeger, "Characterizations of minimal expected length codes", *IEEE Transactions on Information Theory*, (submitted on November 13, 2023). Also available at: arXiv:2311.07007 [cs.IT].

[6] T. M. Cover, "On the competitive optimality of Huffman codes", *IEEE Transactions on Information Theory*, vol. 37, no. , pp. 172 – 174, January 1991.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edition, New Jersey, Wiley-Interscience, 2006.

[8] L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[9] M. Feder, "A note on the competitive optimality of Huffman codes", *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 436 – 439, March 1992.

[10] R. G. Gallager, "Variations on a theme by Huffman", *IEEE Transactions on Information Theory*, vol. 24, no. 6, pp. 668 – 8674, November 1978.

[11] M. Khosravifard, H. Saidi, M. Esmaeili, and T. A. Gulliver, "The minimum average code for finite memoryless monotone sources", *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 955 – 975, March 2007.

[12] T. Linder, V. Tarokh, K. Zeger, "Existence of optimal prefix codes for infinite source alphabets", *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 2026 – 2028, November 1997.

[13] H. B. Newson, "Volume of a polyhedron", *Annals of Mathematics*, vol. 1, no. 1/4, pp. 108 – 110, September 1899.

[14] K. W. Ng, G.-L. Tian, and M.-L. Tang, *Dirichlet and Related Distributions: Theory, Methods and Applications*, John Wiley & Sons, 2011.

[15] P. Rastegari, M. Khosravifard, H. Narimani, and T. A. Gulliver, "On the structure of the minimum average redundancy code for monotone sources", *IEEE Communications Letters*, vol. 18, no. 4, pp. 664 – 667, April 2014.

[16] H. Yamamoto and T. Itoh, "Competitive optimality of source codes", *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 2015 – 2019, November 1995.

[17] H. Yamamoto and H. Yokoo, "Average-sense optimality and competitive optimality for almost instantaneous VF codes", *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2174 – 2184, September 2001.