Existence of Optimal Prefix Codes for Infinite Source Alphabets *

Tamás Linder Vahid Tarokh Kenneth Zeger

Abstract

It is proven that for every random variable with a countably infinite set of outcomes and finite entropy there exists an optimal prefix code which can be constructed from Huffman codes for truncated versions of the random variable, and that the average lengths of any sequence of Huffman codes for the truncated versions converge to that of the optimal code. Also, it is shown that every optimal infinite code achieves Kraft's inequality with equality.

1 Introduction

An *alphabet* $\mathcal{A}$ is a finite set and $\mathcal{A}^*$ is the set of all finite length words formed from the elements of $\mathcal{A}$. For each word $w \in \mathcal{A}^*$, let $l(w)$ denote the word length of w. A D-ary *prefix code* C over an alphabet $\mathcal{A}$ (with $|\mathcal{A}| = D$) is a subset of $\mathcal{A}^*$ with the property that no word in C is the prefix of another word in C. Let $\mathcal{Z}^+$ denote the positive integers.

A sequence of D-ary prefix codes $C_1, C_2, C_3, \ldots$ *converges* to an infinite prefix code C if for every $i \geq 1$, the i^{th} codeword of C_n is eventually constant (as n grows) and equals the i^{th} codeword of C. D-ary prefix codes are known to satisfy Kraft's inequality $\sum_{w \in C} D^{-l(w)} \leq 1$. Conversely any collection of positive integers that satisfies Kraft's inequality corresponds to the codeword lengths of a prefix code.

Let X be a source random variable whose countably infinite range is (without loss of generality) $\mathcal{Z}^+$, with respective probabilities $p_1 \geq p_2 \geq p_3 \geq \cdots$, where $p_i > 0$ for all i. The average length of a prefix code $C = \{w_1, w_2, \ldots\}$ to encode X is $\sum_{i=1}^{\infty} p_i l(w_i)$. A prefix code C is called *optimal* for a source X if no other prefix code has a smaller average length. The entropy of the random variable X is defined as $H(X) = -\sum_{i=1}^{\infty} p_i \log p_i$. It is known that the average length of an optimal prefix code is no smaller than $H(X)$ and is smaller than $H(X) + 1$.

The well-known Huffman algorithm gives a method for constructing optimal prefix codes for sources with finite ranges. For each $n \geq 1$, let X_n be a random variable with a finite range and with outcome probabilities $p_i^{(n)} = p_i / S_n$ for $1 \leq i \leq n$, where

*T. Linder and K. Zeger are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407. T Linder is on leave from the Technical University of Budapest, Hungary. (email: {linder, zeger}@code.ucsd.edu). V. Tarokh is with the AT&T Laboratories, 600 Mountain Avenue Murray Hill, N.J. 07974-0636. (email: tarokh@research.att.com).
This research was supported in part by the National Science Foundation.

$S_n = \sum_{j=1}^{n} p_j$. A *D-ary truncated Huffman cod*
code for X_n.

For sources with infinite ranges, several
optimal codes [1, 2, 3, 4, 5], but in each case so
mass function of the source random variable w
there is no known proof in the literature that o
infinite ranges.

In this correspondence we present such a
particular, we show that a subsequence of Huff
of the source random variable X lead to an op
existence proof and cannot, however, specify wh
However, this theorem does suggest that recurs
might exist for any source, regardless of how fa
decay. We also show that any sequence of trun
the average length sense, whereas only a subs
code sense.

If a source random variable has a finite ran
Kraft's condition with equality, but not neces
contrast, our theorem also establishes that for
source with an infinite range must satisfy the K

In [4] it was noted that an optimal code f
have a full encoding tree. However, a full encod
inequality is satisfied with equality.

A simple counterexample to demonstrate th
$A, B \subset \{0,1\}^*$ let $AB = \{ab \in \{0,1\}^* : a \in A, b$
be the set of all n-bit binary words excluding th
word concatenation. Define the prefix code

$$\begin{aligned} C &= \left(\bigcup_{k=2}^{\infty} \left(\prod_{n=2}^{k} T_n \right) \right. \\ &= \{00, 01000, 100 \end{aligned}$$

and note that the Kraft sum for C is

$$\begin{aligned} \sum_{w \in C} 2^{-l(w)} &= \frac{1}{4} + \sum_{k=2}^{\infty} \left| \left(\prod_{n=} \right. \right. \\ &= \frac{1}{4} + \sum_{k=2}^{\infty} \left(2^{-} \right. \\ &< \frac{1}{4} + \sum_{k=2}^{\infty} 2^{\sum_{i=}^{k}} \\ &= \sum_{k=1}^{\infty} 2^{-(k+1)} \\ &= 1/2. \end{aligned}$$

Thus the Kraft inequality is strict in this case and
of the code C is full.

refix Codes for Infinite Source lphabets *

hid Tarokh Kenneth Zeger

Abstract

ndom variable with a countably infinite set of exists an optimal prefix code which can be con- truncated versions of the random variable, and quence of Huffman codes for the truncated ver- nal code. Also, it is shown that every optimal ality with equality.

s the set of all finite length words formed from the *, let $l(w)$ denote the word length of w. A D-ary n $|\mathcal{A}| = D$) is a subset of $\mathcal{A}^*$ with the property that word in C. Let $\mathcal{Z}^+$ denote the positive integers. $C_1, C_2, C_3, \ldots$ *converges* to an infinite prefix code rd of C_n is eventually constant (as n grows) and prefix codes are known to satisfy Kraft's inequality collection of positive integers that satisfies Kraft's rd lengths of a prefix code.

ble whose countably infinite range is (without loss obabilities $p_1 \geq p_2 \geq p_3 \geq \cdots$, where $p_i > 0$ for all e $C = \{w_1, w_2, \ldots\}$ to encode X is $\sum_{i=1}^{\infty} p_i l(w_i)$. A ource X if no other prefix code has a smaller average ariable X is defined as $H(X) = -\sum_{i=1}^{\infty} p_i \log p_i$. It n optimal prefix code is no smaller than $H(X)$ and

am gives a method for constructing optimal prefix . For each $n \geq 1$, let X_n be a random variable ne probabilities $p_i^{(n)} = p_i/S_n$ for $1 \leq i \leq n$, where

partment of Electrical and Computer Engineering, University 93-0407. T Linder is on leave from the Technical University , zeger}@code.ucsd.edu). V. Tarokh is with the AT&T Hill, N.J. 07974-0636. (email: tarokh@research.att.com).
National Science Foundation.

$S_n = \sum_{j=1}^{n} p_j$. A *D-ary truncated Huffman code* of size n for X is defined to be a Huffman code for X_n.

For sources with infinite ranges, several approaches have been taken to construct optimal codes [1, 2, 3, 4, 5], but in each case some condition on the tail of the probability mass function of the source random variable was assumed. To the best of our knowledge there is no known proof in the literature that optimal codes always exist for sources with infinite ranges.

In this correspondence we present such a proof for sources with finite entropy. In particular, we show that a subsequence of Huffman codes designed for truncated versions of the source random variable X lead to an optimal infinite code for X. We provide an existence proof and cannot, however, specify which Huffman code subsequence is needed. However, this theorem does suggest that recursive Huffman code construction algorithms might exist for any source, regardless of how fast the tails of its probability mass function decay. We also show that any sequence of truncated Huffman codes indeed converges in the average length sense, whereas only a subsequence is guaranteed to converge in the code sense.

If a source random variable has a finite range then an optimal binary code satisfy's Kraft's condition with equality, but not necessarily for D-ary codes when $D \geq 3$. In contrast, our theorem also establishes that for all $D \geq 2$ an optimal D-ary code for a source with an infinite range must satisfy the Kraft inequality with equality.

In [4] it was noted that an optimal code for a source with an infinite range must have a full encoding tree. However, a full encoding tree does not guarantee that Kraft's inequality is satisfied with equality.

A simple counterexample to demonstrate this fact for $D = 2$ is given next. For any $A, B \subset \{0,1\}^*$ let $AB = \{ab \in \{0,1\}^* : a \in A, b \in B\}$. For $n \geq 0$, let $T_n = \{0,1\}^n \setminus \{0^n\}$ be the set of all n-bit binary words excluding the all zeros word, and let $\prod$ denote binary word concatenation. Define the prefix code

$$\begin{aligned} C &= \left(\bigcup_{k=2}^{\infty}\left(\prod_{n=2}^{k} T_n\right)0^{k+1}\right) \cup \{00\} \\ &= \{00, 01000, 10000, 11000, \ldots\} \end{aligned}$$

and note that the Kraft sum for C is

$$\begin{aligned} \sum_{w\in C} 2^{-l(w)} &= \frac{1}{4} + \sum_{k=2}^{\infty}\left|\left(\prod_{n=2}^{k} T_n\right)0^{k+1}\right| 2^{-\sum_{i=2}^{k+1} i} \\ &= \frac{1}{4} + \sum_{k=2}^{\infty}\left(2^{-\sum_{i=2}^{k+1} i}\right)\prod_{n=2}^{k}(2^n - 1) \\ &< \frac{1}{4} + \sum_{k=2}^{\infty} 2^{\sum_{i=2}^{k} i} 2^{-\sum_{i=2}^{k+1} i} \\ &= \sum_{k=1}^{\infty} 2^{-(k+1)} \\ &= 1/2. \end{aligned}$$

Thus the Kraft inequality is strict in this case and it is easy to see that the encoding tree of the code C is full.

2 Main Result

Theorem 1 *Let X be a random variable with a countably infinite set of possible outcomes and with finite entropy. Then for every $D > 1$, the following hold:*

(I) *There exists a sequence of D-ary truncated Huffman codes for X which converges to an optimal code for X.*

(II) *The average codeword lengths in any sequence of D-ary truncated Huffman codes converge to the minimum possible average codeword length for X.*

(III) *Any optimal D-ary prefix code for X must satisfy the Kraft inequality with equality.*

Proof For each $n \geq 1$, let C_n be a D-ary truncated Huffman code of size n for X, and denote the sequence of n codeword lengths of C_n (followed by zeros) by $l^{(n)} = \{l_1^{(n)}, l_2^{(n)}, \ldots, l_n^{(n)}, 0, 0, 0, \cdots\}$. Let $\mathcal{F}$ denote the set of all sequences of positive integers. For each n, the average length $\sum_{i=1}^{\infty} l_i^{(n)} p_i^{(n)}$ of Huffman code C_n is not larger than $H(X_n)+1$, where the entropy of X_n is

$$H(X_n) = -\frac{1}{S_n}\sum_{i=1}^{n} p_i \log p_i - \log\frac{1}{S_n} \to H(X) \text{ as } n \to \infty,$$

since $S_n = \sum_{i=1}^{n} p_i \to 1$ as $n \to \infty$. Hence $H(X_n) + 1 \leq H(X) + 2$, for n sufficiently large, and it is easy to see that $l_i^{(n)} \leq (H(X)+2)/p_i$ for n sufficiently large.

Thus, for each i, the sequence of codeword lengths $\{l_i^{(1)}, l_i^{(2)}, l_i^{(3)}, \cdots\}$ is bounded and therefore the corresponding sequence of codewords can only take on a finite set of possible values. Hence, for each i, there is a convergent subsequence of codewords. We conclude (using a minor modification of [6, Theorem 7.23]) that there exists a subsequence of codes $C_{n_1}, C_{n_2}, C_{n_3}, \ldots$ that converges to an infinite code $\hat{C}$. Clearly $\hat{C}$ is a prefix code since it is a limit of finite Huffman codes. Furthermore the subsequence $\{l^{(n_k)}\}$, of elements of $\mathcal{F}$, converges to a sequence $\hat{l} = \{\hat{l}_1, \hat{l}_2, \cdots\} \in \mathcal{F}$, in the sense that for each $i \in \mathcal{Z}^+$, the sequence $l_i^{(n_k)}$ converges to $\hat{l}_i$.

To show the optimality of $\hat{C}$, let $\lambda_1, \lambda_2, \lambda_3, \ldots$ be the codeword lengths of an arbitrary prefix code. For every k, there exists a j such that $\hat{l}_i = l_i^{(n_m)}$ for every $i \leq k$ provided that $m \geq j$. Thus for all $m \geq j$ and $k \leq n_m$, the optimality of Huffman codes implies

$$\sum_{i=1}^{k} p_i \hat{l}_i = \sum_{i=1}^{k} p_i l_i^{(n_m)} \leq \sum_{i=1}^{n_m} p_i l_i^{(n_m)} = S_{n_m} \sum_{i=1}^{n_m} p_i^{(n_m)} l_i^{(n_m)} \leq S_{n_m} \sum_{i=1}^{n_m} p_i^{(n_m)} \lambda_i. \qquad (1)$$

Therefore, by letting first $m \to \infty$ and then $k \to \infty$, we obtain

$$\sum_{i=1}^{\infty} p_i \hat{l}_i \leq \sum_{i=1}^{\infty} p_i \lambda_i.$$

This implies that the infinite code $\hat{C}$ is optimal.

To prove part (II) of the theorem, notice that by the optimality of Huffman codes

$$\sum_{i=1}^{n} p_i l_i^{(n)} = S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} \leq S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n+1)} = \sum_{i=1}^{n} p_i l_i^{(n+1)} \leq \sum_{i=1}^{n+1} p_i l_i^{(n+1)}.$$

The sequence $\sum_{i=1}^{n} p_i l_i^{(n)}$ is thus an incr
$H(X) + 2$ and has a limit L. It follows fr

$$\sum_{i=1}^{\infty} p_i \hat{l}_i \leq \lim_{m\to\infty} \sum_{i=1}^{n_m} p_i$$

Next by the optimality of Huffman co

$$\lim_{n\to\infty} \sum_{i=1}^{n} p_i l_i^{(n)} = \lim_{n\to\infty} S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} \leq \lim_{n-}$$

Thus

$$\lim_{n\to\infty} \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} = \lim_{n\to\infty}$$

This proves the second part of the theorem

Next we prove part (III) of the theorem. be denoted $l_1 \leq l_2 \leq l_3 \leq \cdots$, and assum strict, i.e. $\sum_i D^{-l_i} < 1$. Let $\delta = 1 - \sum_i D^{-}$ such that $D^{-l_i} < \delta$ for all $i \geq k$. Let j be a of integers $\hat{l}_1, \hat{l}_2, \cdots$ such that $\hat{l}_i = l_i$ for all

$$\sum_{i=1}^{\infty} D^{-\hat{l}_i} = \sum_{i=1}^{\infty} D^{-l_i} - D^{-l_j} +$$

Thus the integers $\hat{l}_1, \hat{l}_2, \cdots$ satisfy Kraft's i having them as codeword lengths. Since $\hat{l}_j$ smaller average codeword length for X tha are $l_1, l_2, \cdots$. This is a contradiction.

References

[1] J. Abrahams, "Huffman-type codes for *Franklin Institute,* vol. 331B, no. 3, pp.

[2] R. A. Gallager and D. C. Van Voorhis, Distributed Integer Alphabets," *IEEE* 228-230, 1975.

[3] P. A. Humblet, "Optimal Source Codin *Trans. Inform. Theory,* vol. 24, no. 1, p

[4] B. Montgomery and J. Abrahams, "On Condition Codes for Finite and Infinite S 33, no. 1, pp. 156-160, 1987.

[5] A. Kato, T. S. Han, and H. Nagaoka, *IEEE Trans. Inform. Theory,* vol. 42, no

[6] W. Rudin, *Principles of Mathematical A* New York, 1976.

e with a countably infinite set of possible outcomes
$D > 1$, the following hold:

uncated Huffman codes for X which converges to

ny sequence of D-ary truncated Huffman codes
average codeword length for X.

X must satisfy the Kraft inequality with equality.

D-ary truncated Huffman code of size n for X,
rd lengths of C_n (followed by zeros) by $l^{(n)} =$
note the set of all sequences of positive integers.
$^{(n)}$ of Huffman code C_n is not larger than $H(X_n)+$

$$p_i - \log \frac{1}{S_n} \to H(X) \text{ as } n \to \infty,$$

lence $H(X_n) + 1 \leq H(X) + 2$, for n sufficiently
$H(X) + 2)/p_i$ for n sufficiently large.
leword lengths $\{l_i^{(1)}, l_i^{(2)}, l_i^{(3)}, \cdots\}$ is bounded and
codewords can only take on a finite set of possible
vergent subsequence of codewords. We conclude
em 7.23]) that there exists a subsequence of codes
nfinite code $\hat{C}$. Clearly $\hat{C}$ is a prefix code since it
thermore the subsequence $\{l^{(n_k)}\}$, of elements of
$\cdots\} \in \mathcal{F}$, in the sense that for each $i \in \mathcal{Z}^+$, the

$\lambda_2, \lambda_3, \ldots$ be the codeword lengths of an arbitrary
a j such that $\hat{l}_i = l_i^{(n_m)}$ for every $i \leq k$ provided
$\leq n_m$, the optimality of Huffman codes implies

$$^{n)} = S_{n_m} \sum_{i=1}^{n_m} p_i^{(n_m)} l_i^{(n_m)} \leq S_{n_m} \sum_{i=1}^{n_m} p_i^{(n_m)} \lambda_i. \qquad (1)$$

then $k \to \infty$, we obtain

$$p_i \hat{l}_i \leq \sum_{i=1}^{\infty} p_i \lambda_i.$$

optimal.
notice that by the optimality of Huffman codes

$$\sum_{1} p_i^{(n)} l_i^{(n+1)} = \sum_{i=1}^{n} p_i l_i^{(n+1)} \leq \sum_{i=1}^{n+1} p_i l_i^{(n+1)}.$$

The sequence $\sum_{i=1}^{n} p_i l_i^{(n)}$ is thus an increasing sequence which is bounded above by $H(X) + 2$ and has a limit L. It follows from (1) that

$$\sum_{i=1}^{\infty} p_i \hat{l}_i \leq \lim_{m \to \infty} \sum_{i=1}^{n_m} p_i l_i^{(n_m)} = \lim_{n \to \infty} \sum_{i=1}^{n} p_i l_i^{(n)}$$

Next by the optimality of Huffman codes

$$\lim_{n \to \infty} \sum_{i=1}^{n} p_i l_i^{(n)} = \lim_{n \to \infty} S_n \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} \leq \lim_{n \to \infty} S_n \sum_{i=1}^{n} p_i^{(n)} \hat{l}_i = \lim_{n \to \infty} \sum_{i=1}^{n} p_i \hat{l}_i = \sum_{i=1}^{\infty} p_i \hat{l}_i.$$

Thus

$$\lim_{n \to \infty} \sum_{i=1}^{n} p_i^{(n)} l_i^{(n)} = \lim_{n \to \infty} \frac{1}{S_n} \sum_{i=1}^{n} p_i l_i^{(n)} = \sum_{i=1}^{\infty} p_i \hat{l}_i.$$

This proves the second part of the theorem.

Next we prove part (III) of the theorem. Let the codeword lengths of an optimal code be denoted $l_1 \leq l_2 \leq l_3 \leq \cdots$, and assume to the contrary that the Kraft inequality is strict, i.e. $\sum_i D^{-l_i} < 1$. Let $\delta = 1 - \sum_i D^{-l_i} > 0$. Then there exists a positive integer k such that $D^{-l_i} < \delta$ for all $i \geq k$. Let j be an integer such that $l_j > l_k$. Define a collection of integers $\hat{l}_1, \hat{l}_2, \cdots$ such that $\hat{l}_i = l_i$ for all $i \neq j$ and such that $\hat{l}_j = l_k$. Then

$$\sum_{i=1}^{\infty} D^{-\hat{l}_i} = \sum_{i=1}^{\infty} D^{-l_i} - D^{-l_j} + D^{-l_k} < \sum_{i=1}^{\infty} D^{-l_i} + \delta = 1.$$

Thus the integers $\hat{l}_1, \hat{l}_2, \cdots$ satisfy Kraft's inequality, so that there exists a prefix code having them as codeword lengths. Since $\hat{l}_j < l_j$, such a prefix code will have a strictly smaller average codeword length for X than the optimal code whose codeword lengths are $l_1, l_2, \cdots$. This is a contradiction. □.

References

[1] J. Abrahams, "Huffman-type codes for infinite source distributions," *Journal of the Franklin Institute,* vol. 331B, no. 3, pp. 265-271, 1994.

[2] R. A. Gallager and D. C. Van Voorhis, "Optimal Source Coding for Geometrically Distributed Integer Alphabets," *IEEE Trans. Inform. Theory,* vol. 21, no. 2, pp. 228-230, 1975.

[3] P. A. Humblet, "Optimal Source Coding for a Class of Integer Alphabets," *IEEE Trans. Inform. Theory,* vol. 24, no. 1, pp. 110-112, 1978.

[4] B. Montgomery and J. Abrahams, "On the Redundancy of Optimal Binary Prefix-Condition Codes for Finite and Infinite Sources," *IEEE Trans. Inform. Theory,* vol. 33, no. 1, pp. 156-160, 1987.

[5] A. Kato, T. S. Han, and H. Nagaoka, "Huffman Coding with Infinite Alphabet," *IEEE Trans. Inform. Theory,* vol. 42, no. 3, pp. 977-984, May 1996.

[6] W. Rudin, *Principles of Mathematical Analysis,* Third Edition, McGraw-Hill Co., New York, 1976.