A straightforward simulation technique is adequate for error probabilities greater than 10^{-4}, but requires too much time to practically evaluate smaller error probabilities. To obtain results for smaller error probabilities, we used importance sampling [7], [8].

In Fig. 3 we show the results of the numerical simulations for a channel with 32-ary orthogonal signaling. The block length of the code was 31 symbols, and the minimum distance of the code was set equal to 7. Although this set of code parameters is the same as those used in Section IV-A, the actual code is quite different, due to the fact that the symbols are no longer binary. In addition, note that the results for this chapter are compared to errors-only coding, and not to GMD decoding. Results for GMD decoding are considerably more difficult to obtain by simulation because the $\boldsymbol{\alpha}_i$'s do not have a convenient distribution. A straightforward simulation for GMD decoding would be slower due to the need to simulate the demodulator output for each letter a_j. In addition, the implementation for improved GMD decoding is nearly the same as that of GMD decoding. Since the improved version is always better, there is no reason why it should not be used instead.

If we compare these results to those obtained for binary orthogonal signals, we can see that the relative performance of improved GMD decoding and errors-only decoding has the same appearance as a function of the signal-to-noise ratio. In fact, the performance difference between improved GMD decoding and errors-only decoding for the results of Section III and this section are almost identical as a function of codeword error probability.

V. Conclusion

In this correspondence, we presented two improvements to the generalized-minimum-distance decoding acceptance criterion. The definition of the reliabilities has been extended so that nonbinary signal sets can be better handled, in particular, it is possible to use the true likelihood metric. In addition, we have developed a new acceptance criterion using the vector reliabilities that is less stringent than previous conditions. We have shown that the performance (when using the new acceptance criterion) of the improved algorithm (in additive-white-Gaussian noise) is asymptotically the same as that of maximum-likelihood decoding for channels using M-ary orthogonal signaling.

References

[1] R. E. Blahut, *Theory and Practice of Error-Control Codes*. Reading, MA: Addison-Wesley, 1983.

[2] M. B. Pursley, "Packet error probabilities in frequency-hop radio networks—Coping with statistical dependence and noisy side information," *IEEE Global Telecommunications Conference Record*, Houston, TX, vol. 1, pp. 165–170, Dec. 1986.

[3] G. D. Forney Jr., *Concatenated Codes*. Cambridge, MA: MIT Press, 1966, pp. 36–62.

[4] ____, "Generalized minimum distance decoding," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 125–131, Apr. 1966.

[5] C. C. H. Yu and D. J. Costello Jr., "Generalized minimum distance decoding algorithms for Q-ary output channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 238–243, Mar. 1980.

[6] G. Einarsson and C. Sundberg, "A note on soft-decision decoding with successive erasures," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 88–96, Jan. 1976.

[7] K. S. Shanmugam and P. Balban, "A modified Monte Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1916–1924, Nov. 1980.

[8] P. M. Hahn and M. C. Jeruchim, "Developments in the theory and application of importance sampling," *IEEE Trans. Commun.*, vol. COM-35, no. 7, pp. 706–714, July 1987.

On the Competitive Optimality of Huffman Codes

Thomas M. Cover

***Abstract*—Let X be a discrete random variable drawn according to a probability mass function $p(x)$, and suppose $p(x)$, is dyadic, i.e., $\log(1/p(x))$ is an integer for each x. We show that the binary code length assignment**

$$l(x) = \log(1/p(x))$$

dominates any other uniquely decodable assignment $l'(x)$ in expected length in the sense that $El(X) < El'(X)$, indicating optimality in long run performance (which is well known), and competitively dominates $l'(x)$, in the sense that $\Pr\{l(X) < l'(X)\} > \Pr\{l(X) > l'(X)\}$, which indicates l is also optimal in the short run. In general, if p is not dyadic, then $l = \lceil \log 1/p \rceil$ dominates $l'+1$ in expected length and competitively dominates $l'+1$, where l' is any other uniquely decodable code.

***Index Terms*—Huffman codes, Shannon codes, competitive optimality, optimality of Huffman codes, data compression.**

I. Introduction

Flying on Mexican airlines into the United States, one observes two signs on the bulkhead: **No smoking**, and under it, **No fumar**. The other says, **Fasten seat belts**, and under it, **Abrocharse el cinturon**. Note that the "Fasten seat belts" sign is shorter in English than in Spanish, while the reverse is true of the "No smoking" sign. Thus English and Spanish are "competitively" equal for this example—each language is shorter half the time. However, the average number of symbols for these two signs clearly favors English over Spanish. Is it conceivable in general that brief translations are shorter in Spanish more often than they are in English, while long translations are shorter in English than they are in Spanish? Mathematically put, we ask whether it is possible that $\Pr(l_E \geq l_S) \geq 1/2$ while $El_E \leq El_S$, where l_E and l_S are the lengths of the English and Spanish versions.

Here is a coding example where one observes this sort of anomalous ordering. We consider a random variable X that takes on four possible values and we assign the encodings C_E and C_S into binary strings as follows:

$X = 1,$	$2,$	$3,$	4
$p(x) = \frac{1}{4},$	$\frac{1}{4},$	$\frac{1}{4},$	$\frac{1}{4}$
$C_E(x) = 000,$	$001,$	$010,$	011
$l_E(x) = 3,$	$3,$	$3,$	3
$C_S(x) = 00,$	$01,$	$10,$	1111111
$l_S(x) = 2,$	$2,$	$2,$	$7.$

The expected description lengths under each code are

$$El_E(X) = 3; \quad El_S(X) = 3\tfrac{1}{4},$$

while the probability that code C_S is shorter than code C_E is

$$\Pr\{l_E(X) > l_S(X)\} = \tfrac{3}{4}.$$

Manuscript received September 1, 1988. This work was supported in part by the National Science Foundation under Contract NCR 89-14538 and JSEP Contract DAAL03-88-0011. This work was presented at the IEEE International Symposium on Information Theory, Kobe, Japan, June 1988.

The author is with the Departments of Electrical Engineering and Statistics, Stanford University, Durand Building, Room 121, Stanford, CA 94305.

IEEE Log Number 9038867.

Notice how the Spanish word length assignment $l_S(x)$ undercuts the English assignment for $x = 1,2,3$. One notes that the expected value of l_E is less than the expected value of l_S. On the other hand, because l_E is dominated by l_S in three out of the four cases, the probability that $l_E > l_S$ is $\frac{3}{4}$. Thus, in this example, (binary) English is longer most of the time but is shorter on the average.

This coding example illustrates the possibility of different orderings under the two criteria, but lacks charm because both encodings are extraordinarily wasteful. There is a reason for this which will be proved in Theorem 1. Apparently optimal codes (Huffman codes, for dyadic distributions) enjoy the distinction of being shorter on the average and also on the average shorter in a sense that will be made precise.

We first review the well understood notion of expected length optimality and then define competitive optimality. An inequality will be proved that will be used to show that Huffman codes for dyadic sources are strictly competitively optimal and strictly expected length optimal. A similar but somewhat weaker result will be proved for nondyadic distributions.

The main point to be made from all of this is that Huffman coding for dyadic distributions has an unexpected bonus. Not only is it expected length optimal, but it cannot be undercut by another code more than half the time, even if the other code is granted infinite expected length.

II. Definitions

We wish to show that codes with word lengths $l(x) = \log 1/p(x)$ are shorter than any other code assignment $l'(x)$ more often than not in the sense that

$$\Pr\{l < l'\} > \Pr\{l > l'\},$$

or equivalently,

$$\sum_x p(x)\,\mathrm{sgn}\,(l(x) - l'(x)) < 0,$$

for all uniquely decodable assignments $l'(x)$, where $\mathrm{sgn}(t)$ is defined by

$$\mathrm{sgn}\,(t) = \begin{cases} 1, & t > 0 \\ 0, & t = 0 \\ -1, & t < 0. \end{cases}$$

Throughout, log denotes log to the base 2.

We recall the theory of data compression for the discrete random variable $X \sim p(x)$, $x \in X = \{x_1, x_2, \cdots, x_m\}$, $p(x_i) > 0$, $\sum p(x_i) = 1$. Let $l(x)$ denote the length of the binary codeword assigned to $x \in X$. By the Kraft–McMillan inequality [4], the word lengths $l(x)$ correspond to a uniquely decodable binary code if and only if

$$\sum_{x \in X} 2^{-l(x)} \le 1. \tag{1}$$

We use the following definitions:

Definition: The probability mass function $p(x)$ is said to be *dyadic* if $\log(1/p(x))$ is an integer for each $x \in X$.

Definition: A code with length assignment l *dominates* code l' *in expected length* if

$$El(X) \le El'(X).$$

Definition: A code l *competitively dominates* l' if

$$\Pr\{l(X) < l'(X)\} \ge \Pr\{l(X) > l'(X)\}.$$

We will say that l is *competitively optimal* if l competitively dominates all other uniquely decodable assignments l'.

Remark: It is worth noting that expected length optimality is not well defined if $H(x) = \infty$, while competitive optimality may still be achievable.

It is known that $l(X) = \lceil \log 1/p(x) \rceil$ codes are close to optimal in expected length, where $\lceil t \rceil$ denotes the least integer $\ge t$, as shown in the following theorem.

Theorem 1 (Shannon [1]): Let $l(x) = \lceil \log 1/p(x) \rceil$. Then

$$H(X) \le El(X) < H(X) + 1 \tag{2}$$

with equality iff $p(x)$ is dyadic. Moreover, if p is dyadic,

$$El(X) \le El'(X), \quad \text{for all } l', \tag{3}$$

if p is dyadic. Finally,

$$El(X) \le E(l'(X) + 1), \quad \text{for all } l'. \tag{4}$$

for any $p(x)$.

Thus l is expected length optimal if p is dyadic and within one of optimal in general.

Proof: By definition of $l(x)$,

$$\log \frac{1}{p(x)} \le l(x) < \log \frac{1}{p(x)} + 1.$$

Taking expectations yields (2). Since any uniquely decodable code has word length assignments $l'(x)$ satisfying (1), the information inequality $\sum p(x) \log p(x)/2^{-l'(x)} \ge 0$ yields $El'(X) \ge H(X)$, with equality iff $l'(x) = l(x)$, thus proving (3). This inequality together with (2) yields (4). □

III. Competitive Optimality

We now examine the performance of the Shannon code

$$l(x) = \lceil \log(1/p(x)) \rceil$$

with respect to the competitive shortness criterion

$$E\,\mathrm{sgn}\,(l'(x) - l(x)).$$

Our proof will be based on the inequality

$$\mathrm{sgn}\,(t) \le 2^t - 1, \quad t = 0, \pm 1, \pm 2, \cdots. \tag{5}$$

Note that this inequality is false if t is unrestricted. We first examine the case where the Shannon code is the Huffman code, which occurs when $p(x)$ is dyadic.

Theorem 2: If p is dyadic, then

$$E\,\mathrm{sgn}\,(l'(x) - l(x)) < 0, \tag{6}$$

for all $l' \ne l$ satisfying the Kraft inequality. This is equivalent to

$$\Pr\{l < l'\} > \Pr\{l > l'\}$$

for all uniquely decodable codes $l' \ne l$.

Proof: Let $l(x) = \log(1/p(x))$. Then

$$\begin{aligned} \Pr\{l > l'\} - \Pr\{l < l'\} &= \sum p(x)\,\mathrm{sgn}\,(l(x) - l'(x)) \\ &\le \sum p(x)(2^{(l(x) - l'(x))} - 1) \\ &= \sum 2^{-l}(2^{l - l'} - 1) \\ &= \sum 2^{-l'} - \sum 2^{-l} \\ &= \sum 2^{-l'} - 1 \\ &\le 0, \end{aligned} \tag{7}$$

where the first inequality follows from (5) and the second from the Kraft inequality. This establishes weak inequality in (6). To show the strict inequality and thus that l is uniquely optimal, we note that the first inequality in (7) is an equality only if $t = 0$ or 1. Thus either $l'(x) = l(x)$, or $l'(x) = l(x) + 1$. If $l'(x) = l(x) + 1$ for any x, then the Kraft inequality is strict: $\sum 2^{l'(x)} < 1$, and (7) is a strict inequality. We conclude that equality holds in (7) if and only if $l'(x) = l(x)$ for all x. Thus l is uniquely optimal. □

IV. Nondyadic p

The Shannon code $\lceil \log 1/p(x) \rceil$ is expected length optimal (a Huffman code) if p is dyadic and is within 1 of expected length optimal for arbitrary p. Similarly, $\lceil \log 1/p(x) \rceil$ is competitively optimal if p is dyadic. We now ask about the competitive performance of $\lceil \log 1/p \rceil$ in general.

Let $l(x) = \lceil \log 1/p(x) \rceil$. We now show that l competitively dominates $l' + 1$ for all uniquely decodable codes l'.

Theorem 3: If $l(x) = \lceil \log(1/p(x)) \rceil$, then

$$E \operatorname{sgn}(l(X) - (l'(X) + 1)) \le 0,$$

for all uniquely decodable assignments $l'(x)$.

Proof: From $l = \lceil \log(1/p) \rceil$ we have $2^{-l} \le p < 2 \cdot 2^{-l}$. Thus

$$\begin{aligned} E \operatorname{sgn}(l(X) - (l'(X) + 1)) &\le \sum p(x)(2^{l(x) - l'(x) - 1} - 1) \\ &= \frac{1}{2} \sum p(x) 2^{l(x) - l'(x)} - 1 \\ &< \frac{2}{2} \sum 2^{-l}(2^{l - l'}) - 1 \\ &= \sum 2^{-l'} - 1 \le 0. \end{aligned}$$
□

V. Repeated Plays

The previous results easily extend to sequences of random variables. Suppose $p(x)$ is dyadic and we wish to encode blocks $(X_1, X_2, \cdots, X_n)$, where $X_1, X_2, \cdots, X_n$ are independent identically distributed according to $p(x)$. Consider the myopic encoding $l(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} l(X_i)$, where $l(x_i) = \log 1/p(x_i)$, obtained by concatenating the codewords associated with the individual symbols.

We observe that $p(x_1, \cdots, x_n)$ is also dyadic, and $l(x_1, \cdots, x_n) = \log(1/p(x_1, \cdots, x_n))$. Consequently,

$$E \operatorname{sgn}(l'(X_1, X_2, \cdots, X_n) - l(X_1, X_2, \cdots, X_n)) > 0$$

for all $l' \ne l$, for all n. Thus the short term goal of designing the competitively shortest code at time $n = 1$ is completely compatible with designing the shortest code for any time. Simply concatenate the codewords.

VI. Summary

Let $l(x) = \lceil \log(1/p(x)) \rceil$. Then for any other uniquely decodable assignment $l'(x)$ we have shown that l competitively dominates $l' + 1$ and also dominates $l' + 1$ in expected value. If p is dyadic, l competitively dominates l' and also dominates l' in expected value. These results indicate that the Shannon codeword length assignment $l(x) = \lceil \log(1/p(x)) \rceil$ has optimal short run as well as optimal long run properties.

Acknowledgment

The author would like to thank Laura Ekroot for helpful discussions.

References

[1] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[2] R. Blahut, *Principles and Practice of Information Theory*. New York: Addison Wesley, 1987.

A Note on D-ary Huffman Codes

Renato M. Capocelli, *Senior Member, IEEE* and
Alfredo De Santis, *Member, IEEE*

***Abstract*—An upper bound on the redundancy of D-ary Huffman codes in terms of the probability p_1 of the most likely source letter is provided. For large values of p_1, the bound improves the one given by Gallager. Additionally, some results known for the binary case ($D = 2$) are extended to arbitrary D-ary Huffman codes. As a consequence, a tight lower bound that corrects a bound recently proposed by Golic and Obradovic is derived.**

***Index Terms*—Huffman coding, entropy, codeword, source coding.**

I. Introduction

Let A be a discrete source with N letters, $2 \le N < \infty$, and p_k denote the probability of letter a_k, $1 \le k \le N$. Let D, $2 \le D < \infty$, denote the size of the code alphabet. Let $\{x_1, x_2, \cdots, x_N\}$ be a set of D-ary codewords and $n_1, n_2, \cdots, n_N$ be the codeword lengths. The *Huffman* encoding algorithm provides an *optimal prefix code* C for the source A. The encoding is optimal in the sense that codeword lengths minimize the *redundancy* r, defined as the difference between the *average codeword length* E of the code and the *entropy* $H(p_1, p_2, \cdots, p_N)$ of the source:

$$r = E - H(p_1, p_2, \cdots, p_N) = \sum_{i=1}^{N} p_i n_i + \sum_{i=1}^{N} p_i \log_D p_i.$$

According to Shannon's first theorem, the redundancy of any Huffman code is always nonnegative and less than or equal to one.

In a remarkable paper [1], Gallager has proved that, knowing the probability of the most likely source letter p_1, the following upper bound holds:

$$r \le \sigma_D + p_1 D / \ln D, \tag{1}$$

where $\sigma_D = \log_D(D-1) + \log_D(\log_D e) - \log_D e + (D-1)^{-1}$. For binary codes ($D = 2$) bounds better than (1) are known [1], [3], [4], [5], and [6]. Bound (1) improves the Shannon limit, $r \le 1$, only when $p_1 < \gamma_D = (1 - \sigma_D)(\ln D)/D$. Moreover γ_D approaches 0 as D gets large. Indicatively, one has that $\gamma_3 = 0.316$, $\gamma_5 = 0.259$, $\gamma_{10} = 0.168$ and $\gamma_{20} = 0.099$. Finding upper bounds tighter than the Shannon limit for $\gamma_D < p_1 < 1$ is therefore an open problem.

A necessary and sufficient condition for the most likely letter of a discrete source to be coded by a single symbol with a binary Huffman code was first obtained by Johnsen [3]. Capocelli *et al.* [4] extended this result to the case of a two symbol codeword.

Manuscript received April 4, 1988; revised August 7, 1989. This work was supported in part by the Italian Ministry of Education and by the Italian National Council for Research. This work was presented in part at IEEE 1988 International Symposium on Information Theory, Kobe, Japan, June 1988.

R. M. Capocelli is with the Dipartimento di Matematica, Università di Roma, 00185 Roma, Italy.

A. Desantis was at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, while on leave from the Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (Salerno), Italy.

IEEE Log Number 9039284.

0018-9448/91/0100-0174$01.00 © 1991 IEEE