Asymptotic Average Redundancy of Huffman (and Other) Block Codes*

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Abstract

We study the redundancy of Huffman (and other) codes. It has been known from the inception of the Huffman code that in the worst case its redundancy — defined as the excess of the code length over the optimal (ideal) code length — is not more than one. Over more than forty years, insightful, elegant and useful constructions have been set up to determine tighter bounds on the Huffman code redundancy. However, to the best of our knowledge no precise asymptotic results have been reported in literature thus far. We consider here a memoryless binary source generating a sequence of length n distributed as $Binomial(n, p)$ with p being the probability of emitting 0. Based on recent results of Stubley, we prove that for $p < 1/2$ the average redundancy $\bar{R}_n^H$ of the Huffman code becomes as $n \to \infty$

$$\bar{R}_n^H = \begin{cases} \frac{3}{2} - \frac{1}{\ln 2} + o(1) = 0.057304\ldots, & \alpha \quad \text{irrational} \\ \frac{3}{2} - \frac{1}{M}\left(\langle \beta M n \rangle - \frac{1}{2}\right) - \frac{1}{M(1-2^{-1/M})} 2^{-\langle n\beta M\rangle/M} + O(\rho^n), & \alpha = \frac{N}{M} \quad \text{rational} \end{cases}$$

where $\alpha = \log_2(1-p)/p$ and $\beta = -\log_2(1-p)$, $\rho < 1$, M, N are integers such that $\gcd(N, M) = 1$, and $\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of x. The appearance of the fractal-like function $\langle \beta M n \rangle$ explains the erratic behavior of the Huffman redundancy, and its "resistance" to succumb to a precise analysis. As a side result, we prove that the average redundancy of the Shannon block code is as $n \to \infty$

$$\bar{R}_n^S = \begin{cases} \frac{1}{2} + o(1) & \alpha = \log_2(1-p)/p \quad \text{irrational} \\ \frac{1}{2} - \frac{1}{M}\left(\langle M n \beta \rangle - \frac{1}{2}\right) + O(\rho^n) & \alpha = \frac{N}{M} \quad \text{rational} \end{cases}$$

where $\rho < 1$. Finally, we derive the redundancy of the Golomb code (for the geometric distribution) which can be viewed as a special case of the Huffman code for infinite alphabets. Unlike the redundancy of Huffman and Shannon codes, Golomb's code redundancy has only oscillating behavior (i.e., there is no convergent mode). These findings are obtained by analytic methods such as theory of distribution of sequences modulo 1 and Fourier series. These analytic tools are able to handle redundancy of a larger class of codes such as extensions of Golomb's code, Krichevsky–Trofimov universal coding, Context–Tree Weighting coding method, minimax universal codes, etc.

Index Terms — Huffman code, Shannon–Fano code, Golomb code, average redundancy, Fourier analysis, distribution of sequences modulo 1, Weyl's criterion.

*The material in this paper was presented in part at the *IEEE International Symposium on Information Theory*, Sorrento, June 25-30, 2000.

†This research was supported in part by NSF Grants NCR-9415491 and C-CR-9804760, and Purdue Faculty Grant GIFC-9919.

1 Introduction

Since the appearance more than 45 years ago of Huffman's classical paper [13] on optimal variable length source coding, Huffman coding still remains one of the most familiar topics in information theory (cf. [1, 2, 3, 9, 18, 25, 26]). Recall that Huffman code is a recursive algorithm built over the associated Huffman tree, in which the two nodes with lowest weights are combined into a new node with a weight that is the sum of the weights of its two children. While Huffman has already known that the average code length is asymptotically equal to the entropy of the source, the asymptotic performance of the code is still not fully understood. Here, for the first time we present asymptotically precise results for the average redundancy of Huffman (and other related) codes. Before we describe our findings, we briefly discuss the redundancy rate problem for lossless coding.

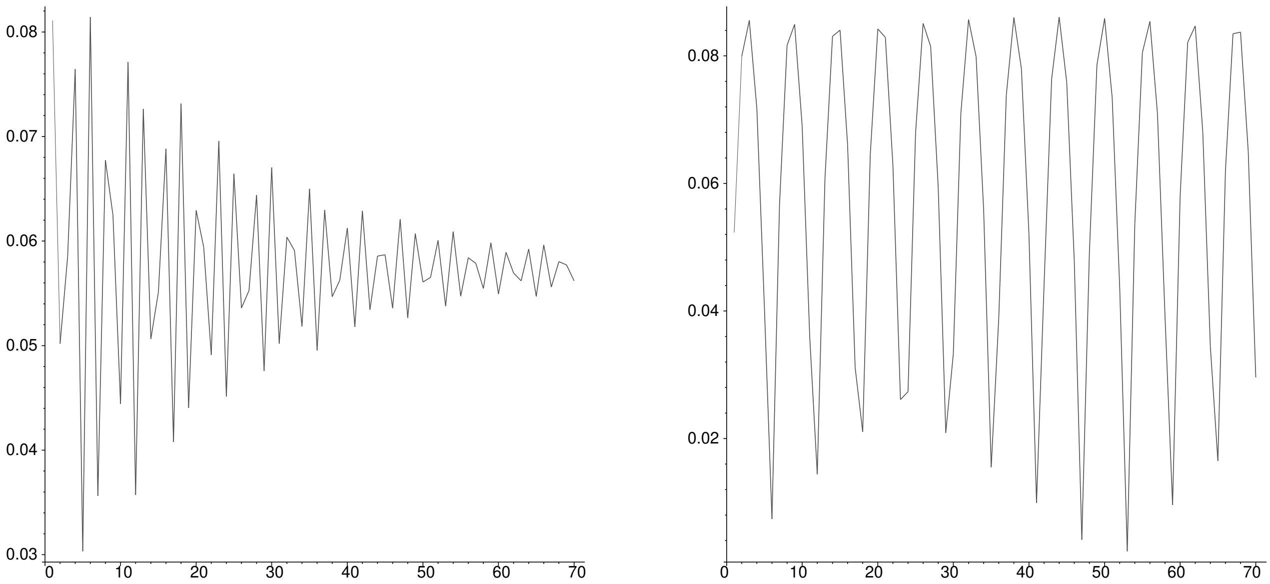
Recent years have seen a resurgence of interest in redundancy rates of lossless coding; (cf. [9, 14, 17, 21, 23, 22, 24, 25, 27, 29]. The redundancy rate problem for a class of sources corresponds to determining how much the actual code length exceeds the optimal (ideal) code length. We define a code $C_n : \mathcal{A}^n \to \{0,1\}^*$ as a mapping from the set $\mathcal{A}^n$ of all sequences of length n over the alphabet $\mathcal{A}$ to the set $\{0,1\}^*$ of binary sequences. We write X_1^n to denote the random variable representing a message of length n. Given a probabilistic source model and a code C_n, we let $P(x_1^n)$ be the probability of the message $x_1^n \in \mathcal{A}^n$, and $L(C_n, x_1^n)$ be the code length for x_1^n and code C_n. Information-theoretic quantities are expressed in binary logarithms written $\lg := \log_2$.

From Shannon's works we know that the entropy $H_n(P) = -\sum_{x_1^n} P(x_1^n) \lg P(x_1^n)$ is an absolute lower bound on the expected code length. Hence $-\lg P(x_1^n)$ can be viewed as the "ideal" code length. The *pointwise redundancy* $R_n(C_n, P; x_1^n)$ and the *average redundancy* $\bar{R}_n(C_n, P)$ are defined as

$$
\begin{aligned}
R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \lg P(x_1^n) \\
\bar{R}_n(C_n, P) &= \mathbf{E}_{X_1^n}[R_n(C_n, P; X_1^n)] = \mathbf{E}[L(C_n, X_1^n)] - H_n(P)
\end{aligned}
$$

where the underlying probability measure P represents a particular source model and $\mathbf{E}$ denotes the expectation.

In this paper, we shall deal with the average redundancy of certain codes (Huffman, Shannon, Golomb, etc) for which the underlying distribution is *known*. We first concentrate on Huffman block codes of length n. A survey on recent results on the redundancy of Huffman codes can be found in [1], so we only address it briefly. Many elegant, insightful and useful lower and upper bounds on the average redundancy of Huffman codes, $\bar{R}_n^H$, are known. Gallager [9] proved that $\bar{R}_n^H \le p_1 + \lg(2(\lg e)/e) \approx p_1 + 0.086$ where p_1 is the probability of the most likely



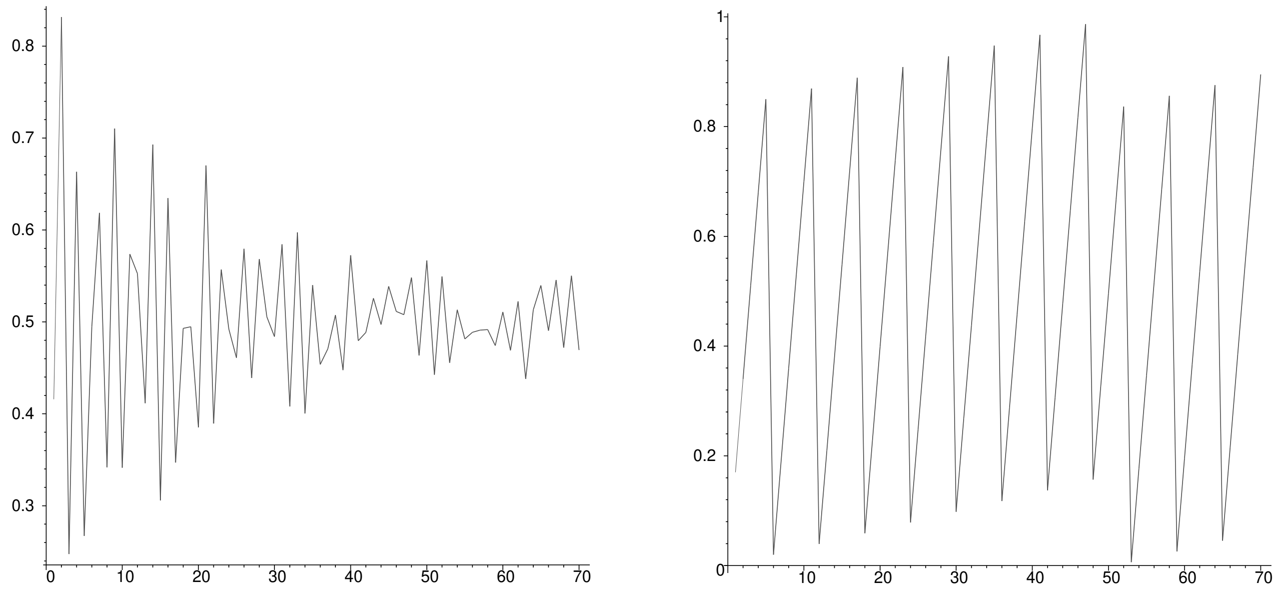
(a) (b)

Figure 1: The average redundancy of Huffman codes versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

symbol. This bound was further improved by Capocelli and de Santis [2, 3], Manstetten [18], Stubley [25, 26]. To the best of our knowledge, no asymptotic results have been reported in literature on the average redundancy of Huffman codes.

To understand the level of difficulty in establishing such asymptotics, we use the tight bounds of Stubley [25]. In Figure 1 we plot the average redundancy $\bar{R}_n^H$ of the Huffman block code for a binary sequence of length n emitted by a memoryless source (where p is the probability of generating 0). In Figure 1(a) we consider $\alpha = \lg(1-p)/p$ irrational while in Figure 1(b) α is rational. The distinct characteristics of these two curves are clearly visible. The function in Figure 1(a) seems to converge to a constant (≈ 0.05) for large n, while the curve in Figure 1(b) is quite erratic (with the maximum close to Gallager's upper bound 0.086!). We shall theoretically justify this behavior in Theorem 1 of Section 2 (cf. abstract for a brief summary of main theoretical findings).

The erratic behavior of the redundancy seems to be a rule rather than an exception. We have already observed this in the redundancy of the Lempel-Ziv and other codes (cf. [14, 17, 22]). Actually, one does not need to look too far since the simplest code, that of Shannon, exhibits the same kind of behavior. In this case, the average redundancy $\bar{R}_n^S$ can be computed



(a) (b)

Figure 2: The average redundancy of Shannon codes versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

as

$$\bar{R}_n^S = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \left(\lceil -\log_2(p^k(1-p)^{n-k}) \rceil + \log_2(p^k(1-p)^{n-k}) \right)$$

provided the underlying probability is $Binomial(n, p)$. In Figure 2 we plot $\bar{R}_n^S$ for $\alpha = \lg(1-p)/p$ irrational and rational. Again, the curves exhibit two different modes. In Theorem 2 of Section 2 we provide a theoretical justification for this behavior. Our main result concerning the redundancy of Shannon codes is also reported in the abstract.

One may wonder whether these two modes, namely, convergent and oscillating, occur in redundancy of other codes with known probability distribution. This turns out not to be the case as shown in Figure 3 where we plot the redundancy of the Golomb code for $Geometric(\theta)$ distribution as $\ell \to \infty$ where $\theta^{-\ell} = 2$. In Theorem 3 of Section 2 we derive the redundancy of the Golomb code and prove that it possesses only the oscillating mode.

The main analytic tool used in the derivation of our results is theory of *uniform* and *Bernoulli distributed sequences modulo 1* (cf. [7, 16, 28]). This technique seems to have other applications in information theory (e.g., redundancy of extensions of Golomb's code [10, 19], KT-estimator [30], the redundancy of the Context–Tree Weighting scheme [30], and universal minimax coding [24]). This theory was already applied by Gray [12] in analyzing

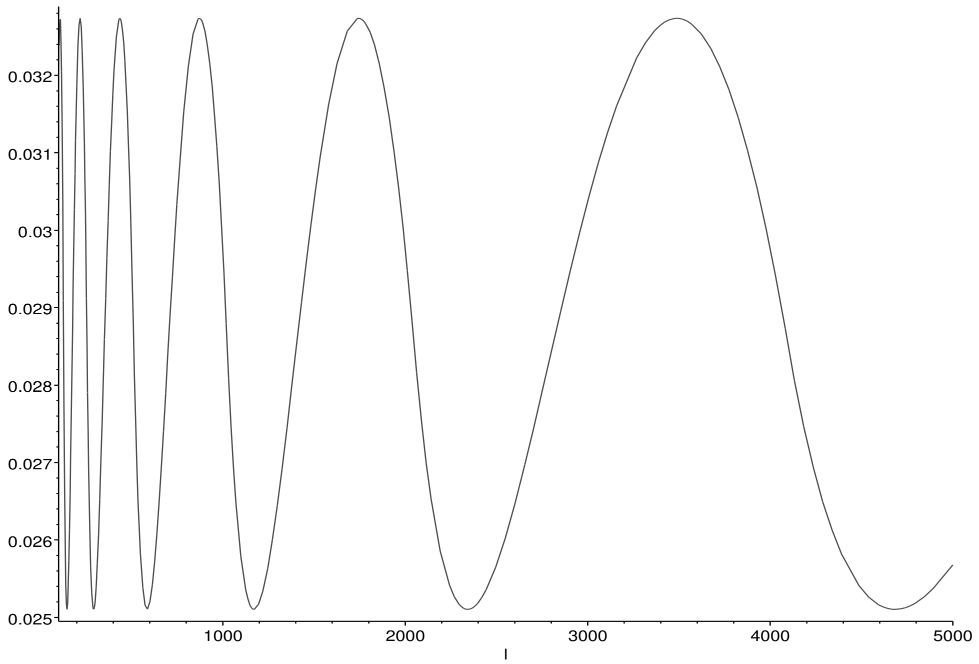


Figure 3: The average redundancy of Golomb codes versus ℓ.

uniform quantizers, where uniform distributed sequences modulo 1 were used to estimate the quantization error.

The rest of the paper is organized as follows. In the next section we present our main results for Huffman, Shannon and Golomb codes (cf. Theorem 1, Theorem 2 and Theorem 3). We prove them in the next three sections. In Section 3 we discuss the redundancy of the Shannon (and Shannon–Fano) code, while in Section 4 we derive the formula for the Huffman code redundancy. Finally, in Section 5 we obtain the asymptotic expansion for the redundancy of the Golomb code.

2 Main Results

We start with a brief description of Stubley's [25] tight bounds on the Huffman code redundancy for binomially distributed blocks of length n. Let p denote the probability of generating "0" and $q = 1 - p$ denote the probability of emitting "1". Throughout, we assume that $p \neq \frac{1}{2}$, however, to simplify some derivations we set $p < \frac{1}{2}$. Certainly, this does not restrict the generality of the analysis. We also write $p(k) = p^k(1-p)^{n-k} = p^k q^{n-k}$.

Let $\mathbf{E}[L_n]$ denote the average length of the Huffman code that can be written as

$$\mathbf{E}[L_n] = \sum_{k=0}^{n} \binom{n}{k} p(k) L(k), \qquad (1)$$

where

$$L(k) = \frac{1}{\binom{n}{k}} \sum_{j \in \mathcal{S}_k} l_j \tag{2}$$

with $\mathcal{S}_k$ representing the set of all inputs having probability $p(k)$, and l_j being the length of the jth code. By Gallager's sibling property, we know that code lengths in $\mathcal{S}_k$ are either equal to $l(k)$ or $l(k)+1$ for some integer $l(k)$. If n_k denotes the number of code words in $\mathcal{S}_k$ that are equal to $l(k)+1$, then

$$L(k) = l(k) + \frac{n_k}{\binom{n}{k}}. \tag{3}$$

Thus, the average redundancy $\bar{R}_n^H$ of the Huffman code is

$$\bar{R}_n^H = \sum_{k=0}^{n} \binom{n}{k} p(k)[L(k) + \lg p(k)] \tag{4}$$

where, we recall, $\lg := \log_2$.

In order to find $l(k)$ and n_k, which are necessary to evaluate (4), one observes that the Huffman code minimizes $\bar{R}_n^H = \mathbf{E}[L_n] - nH$ subject to $\sum_{k=0}^{n} \binom{n}{k} 2^{-L(k)} \leq 1$. Here, $H = -p \lg p - q \lg q$ is the entropy rate. This can be viewed as a simple optimization problem. Using Lagrangian multipliers one can solve it and prove that $L(k) = -\lg p(k)$ if no integer constraints are imposed. If we set $L(k) = l(k) + n_k/\binom{n}{k}$, then one gets

$$l(k) = \lfloor -\lg p(k) \rfloor, \tag{5}$$

$$n_k = \binom{n}{k} [-\lg p(k) - l(k)]\,. \tag{6}$$

where we still did not impose integer constraints on n_k.

Stubley [25] analyzed carefully the above optimization problem and concluded that setting

$$n_k = \left\lceil 2\binom{n}{k} \left(1 - p(k) 2^{l(k)}\right) \right\rceil \tag{7}$$

solves it with integer constraints imposed and with the Kraft inequality being satisfied. After algebraic manipulations, he was led to the following asymptotic formula

$$\bar{R}_n^H \sim \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} f_k + 2 \sum_{k=0}^{n-1} \binom{n}{k} p^k q^{n-k} (1 - 2^{f_k}) \tag{8}$$

where

$$f_k = \lg p(k) + \lfloor -\lg p(k) \rfloor. \tag{9}$$

In the above $a_n \sim b_n$ $\lim_{n\to\infty} a_n/b_n = 1$.

One can re-derive (8) in a slightly stronger form. Let for $p < \frac{1}{2}$

$$\begin{aligned}\alpha &= \log_2\left(\frac{1-p}{p}\right),\\ \beta &= \log_2\left(\frac{1}{1-p}\right).\end{aligned}$$

Following Stubley's derivations, after some algebra, we arrive at

$$\bar{R}_n^H = 2 - \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\langle \alpha k + \beta n\rangle - 2\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}2^{-\langle \alpha k+\beta n\rangle} + O(\rho^n), \tag{10}$$

where $\rho < 1$ and $\langle x\rangle = x - \lfloor x\rfloor$ being the fractional part of x.

We are now in a position to present our main result concerning the asymptotic behavior of $\bar{R}_n^H$ as $n \to \infty$. The proof can be found in Section 4.

Theorem 1 *Consider the Huffman block code of length n over a binary memoryless source Binomial(n,p). Then, for $p < \frac{1}{2}$ as $n \to \infty$*

$$\bar{R}_n^H = \begin{cases} \frac{3}{2} - \frac{1}{\ln 2} + o(1) \approx 0.057304 & \alpha \quad \text{irrational} \\ \\ \frac{3}{2} - \frac{1}{M}\left(\langle \beta M n\rangle - \frac{1}{2}\right) - \frac{1}{M(1-2^{-1/M})}2^{-\langle n\beta M\rangle/M} + O(\rho^n) & \alpha = \frac{N}{M} \end{cases} \tag{11}$$

where N, M are integers such that $\gcd(N, M) = 1$ *and* $\rho < 1$.

Remark. An extension of Theorem 1 to multi-alphabet is possible. Let us consider a V-ary alphabet $\mathcal{A} = \{1, 2, \ldots, V\}$. Based on Stubley's 1998 paper [26], we conclude that in this case the following expression for the Huffman redundancy holds

$$\begin{aligned}\bar{R}_n^H &= 2 - \sum_{k_1+\cdots+k_V=n}\binom{n}{k_1,\ldots,k_V}p_1^{k_1}\cdots p_V^{k_V}\langle \alpha_1 k_1 + \cdots + \alpha_{V-1}k_{V-1} + \beta n\rangle \\ &- 2\sum_{k_1+\cdots+k_V=n}\binom{n}{k_1,\ldots,k_V}p_1^{k_1}\cdots p_V^{k_V}2^{-\langle \alpha_1 k_1+\cdots+\alpha_{V-1}k_{V-1}+\beta n\rangle} + O(\rho^n),\end{aligned}$$

where $\alpha_i = \lg(p_V/p_i)$ for $1 \le i \le V-1$ and $\beta = \lg(1/p_V)$. Using the same arguments as in the proof of Theorem 1, we observe that if one of α_i is irrational, then the formula for $\bar{R}_n^H$ from Theorem 1 remains valid, while if all α_i are rationals, then one derives an expression similar to the one in Theorem 1. $\square$

One may wonder why the average redundancy $\bar{R}_n^H$ exhibits two different "modes" for α irrational and rational. To see this, we must go back to (10) and inspect the term $\langle \alpha k + \beta n\rangle$ as $k, n \to \infty$. From theory of *sequences uniformly distributed modulo 1* (cf. [7, 16] and Section 3) it is known that $\langle \alpha k + \beta n \rangle$ "fills" up the interval $[0, 1)$ "uniformly" for α irrational as $n \to \infty$ (i.e., it can be viewed as the uniform distribution over $[0, 1]$). A different behavior occurs when α is rational. In this case, the sequence $\langle \alpha k + \beta n \rangle$ is periodic, only certain points in the interval $[0, 1]$ are selected, and this fact is reflected in the final formula on the redundancy.

Comparing our theoretical finding with the numerical result of Figure 1, we observe a good agreement. Indeed, in the irrational case, the redundancy curve in Figure 1(a) converges to a constant that is approximately equal to 0.056 which coincides with the result of Theorem 1. In the rational case, we observe that the redundancy swings from almost zero to about 0.086. To see it more precisely, and in fact to recover Gallager's upper bound, we set $x = \langle Mn\beta \rangle$. We first observe that $M = 1$ maximizes $\bar{R}_n^H$, and then

$$\bar{R}_n^H(x) \sim 2 - x - 2^{-x+1}. \tag{12}$$

This leads to

$$\max_{0 \le x < 1} \bar{R}_n^H \sim 1 - \frac{1 + \ln\ln 2}{\ln 2} = \lg(2(\lg e)/e) = 0.08607\ldots, \tag{13}$$

which is the Gallager upper bound (since the most likely probability $p_1 = O(1/\sqrt{n})$ in this case). We formulate it as a corollary.

Corollary 1 *The maximum value of the average Huffman redundancy is*

$$\max\{\bar{R}_n^H\} = 1 - \frac{1 + \ln\ln 2}{\ln 2} + o(1) = \lg(2(\lg e)/e) + o(1) \sim 0.08607\ldots, \tag{14}$$

as $n \to \infty$.

As mentioned in the introduction, the redundancy of the Huffman code is related to the redundancy of the **Shannon** and **Shannon–Fano** codes that we discuss next. Since the Shannon–Fano code needs only one more bit than the Shannon code, we focus here on the latter. In fact, we shall see that the Huffman redundancy $\bar{R}_n^H$ can be expressed in terms of the average Shannon redundancy $\bar{R}_n^S$ as

$$\bar{R}_n^H = 1 + \bar{R}_n^S - 2\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle}. \tag{15}$$

Indeed, the Shannon code assigns length $\lceil -\lg p(k) \rceil$ to a word of probability $p(k) = p^k q^{n-k}$. Therefore, the average Shannon redundancy is

$$\begin{aligned}
\bar{R}_n^S &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left(\lceil -\lg p(k) \rceil + \lg p(k)\right) \\
&= 1 + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left(\lg p(k) + \lfloor -p(k) \rangle\right) \\
&= 1 - \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle \qquad (16)
\end{aligned}$$

where, as before, $\alpha = \lg((1-p)/p$ and $\beta = \lg(1/(1-p)$.

In Section 3 we prove the following asymptotic performance of the Shannon code. (In passing we should point out that numerical results of Figure 2 coincide with the theoretical findings of Theorem 2.)

Theorem 2 *Consider the Shannon block code of length n over a binary memoryless source Binomial(n,p). Then, for $p < \frac{1}{2}$ as $n \to \infty$*

$$\bar{R}_n = \begin{cases} \frac{1}{2} + o(1) & \alpha \quad \text{irrational} \\ \\ \frac{1}{2} - \frac{1}{M}\left(\langle Mn\beta\rangle - \frac{1}{2}\right) + O(\rho^n) & \alpha = \frac{N}{M}, \quad \gcd(N,M) = 1 \end{cases} \tag{17}$$

where $\rho < 1$.

Remark. The rates of convergence in Theorem 1 and Theorem 2 for the irrational case are not specified. This problem is actually quite subtle, and there is no one simple bound for the rate of convergence for all α. However, for almost all irrational α the $o(1)$ term in Theorem 2 can be replaced by $O\left(\frac{\log^{1+\delta} n}{\sqrt{n}}\right)$ for some $\delta > 0$. □

Finally, we consider Golomb's code that can be viewed as a special case of Huffman's code adapted to infinite alphabets. More precisely, let

$$P(i) = (1-\theta)\theta^i, \qquad i \in \mathbb{Z}^+, \tag{18}$$

be the probability assignment on the set of nonnegative integers where $0 < \theta < 1$. Golomb [11] proposed the following optimal binary code (we follow here the description form [10]): Let ℓ be an integer such that

$$\theta^\ell = \frac{1}{2}. \tag{19}$$

In Golomb's code an integer i is represented as $i = \ell j + r$, where $j = \lfloor i/\ell \rfloor$. We encode j by a unary code (i.e., j zeros followed by a one) while r is encoded by a Huffman code (for a uniform distribution on ℓ symbols).

In [10, 11] it is shown that the average Golomb's code length $\mathbf{E}[L]$ is

$$\mathbf{E}[L] = \lceil \log_2 \ell \rceil + \frac{\theta^{2^{\lfloor \log_2 \ell \rfloor + 1}}}{\theta^\ell - 1}. \tag{20}$$

Roughly speaking, the first term above corresponds to the Huffman code while the second term represents the unary coding. Thus, the average redundancy $\bar{R}_\ell^G$ of the Golomb code becomes

$$\bar{R}_\ell^G = \mathbf{E}[L] - H(\theta) \tag{21}$$

where the entropy $H(\theta)$ of the geometric distribution (18) can be computed as

$$H(\theta) = -\log_2(1-\theta) - \frac{\theta}{1-\theta}\log_2\theta. \tag{22}$$

It is of interest to estimate $\bar{R}_\ell^G$ as $l \to \infty$ or equivalently as $\theta \to 1$ (cf. [10, 19]).

In Section 5 we prove the following theorem. One should observe excellent accuracy between the below theoretical findings and the numerical results presented in Figure 3.

Theorem 3 *Consider the Golomb code over an infinite alphabet generated by a source Geometric(θ). Then as $\theta \to 1$ or equivalently $\ell \to \infty$*

$$\bar{R}_\ell^G = 1 - \langle \log_2 \ell \rangle + 4 \cdot 2^{-2^{1-\langle \log_2 \ell \rangle}} - \log_2(\log_2 e) - \log_2 e - \frac{1}{2\ell} + O(\ell^{-2}) \tag{23}$$

where, as before, $\langle x \rangle$ denotes the fractional part of x. Furthermore, the average redundancy $\bar{R}_\ell^G$ oscillates around $2 - \log_2(\log_2 e) - \log_2 e = 0.028538562\ldots$ with

$$\min\{\bar{R}_\ell^G\} = 0.0251005712\ldots, \tag{24}$$

$$\max\{\bar{R}_\ell^G\} = 0.0327344112\ldots \tag{25}$$

as $\ell \to \infty$.

In passing we mention that Theorem 3 can be extended to Golomb's code discussed in [10] where the restriction (19) was relaxed, and its extension recently proposed in [19]. Surprisingly enough, one can prove that (23) still holds for the extension discussed in [10].

3 Derivation of the Redundancy for Shannon Code

We derive our main results in the reverse order we presented them. First, we deal with the redundancy of the Shannon code establishing Theorem 2, and later in Section 4 we extend the analysis to the Huffman code redundancy. This is a natural order of showing our results since according to (15) $\bar{R}_n^S$ is part of $\bar{R}_n^H$. Finally, in Section 5 we prove Theorem 3 concerning the redundancy of the Golomb code.

We are interested in the asymptotics of $\bar{R}_n^S$ given by (16). We start with a Fourier series of $\langle x \rangle$, namely (cf. [12, 32]), for $x \in \mathbb{R}$

$$\begin{aligned}\langle x \rangle &= \frac{1}{2} - \sum_{m=1}^{\infty} \frac{\sin 2\pi m x}{m\pi} \\ &= \frac{1}{2} - \sum_{m \in \mathbb{Z}-\{0\}} c_m e^{2\pi i m x}, \qquad c_m = -\frac{i}{2\pi m},\end{aligned} \tag{26}$$

where $\mathbb{Z}$ is the set of all integers. Hereafter, we shall write $\sum_{m\neq 0} := \sum_{m\in\mathbb{Z}-\{0\}}$.

3.1 Irrational Case

We first deal with the case when α is irrational. Using (26) in (16) we obtain

$$\begin{aligned}\bar{R}_n^S &= \frac{1}{2} + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \sum_{m\neq 0} c_m e^{2\pi i m(\alpha k+\beta n)} \\ &= \frac{1}{2} + \sum_{m\neq 0} c_m e^{2\pi i m\beta n} \left(p e^{2\pi i m\alpha} + q\right)^n . \end{aligned} \tag{27}$$

Our goal now is to prove that the last sum in the above is $o(1)$ when α is irrational. Unfortunately, this seems to be difficult to accomplish within the Fourier series framework. We turn our attention to theory of *sequences distributed modulo 1* (cf. [7, 16]).

We now briefly describe elements of theory of sequences distributed modulo 1 that fits our needs and finds other applications in information theory (cf. [12]). We start with the following definition.

Definition 1 (B-u.d. mod 1) *A sequence $x_n \in \mathbb{R}$ is said to be Bernoulli uniformly distributed modulo 1 (in short: B-u.d. mod 1) if for fixed* $0 < p < 1$

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \chi_I(\langle x_k \rangle) = \lambda(I) \tag{28}$$

holds for every interval $I \subset \mathbb{R}$, where $\chi_I(x_n)$ is the characteristic function of I (i.e., it equals to 1 *if $x_n \in I$ and* 0 *otherwise) and $\lambda(I)$ is the Lebesgue measure of I.*

Remark. If we replace in (28) the binomial distribution by the uniform distribution, then we define the *uniform distributed* sequences modulo 1, or in short u.d. mod 1. Not surprisingly, the property of $\langle x_k \rangle$ does not change when the uniform weight is replaced by the binomial weight since $\langle x_k \rangle$ stills "fills" up densely the interval $(0, 1)$ (cf. [7] for more details). □

The following result summarizes the main property of B-u.d. modulo 1 sequences that we need in the analysis.

Theorem 4 *Let $0 < p < 1$ be a fixed real number and suppose that the sequence x_n is B-uniformly distributed modulo 1. Then for every Riemann integrable function $f : [0, 1] \to \mathbb{R}$ we have*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle) = \int_0^1 f(t)\, dt, \tag{29}$$

where the convergence is uniform for all shifts $y \in \mathbb{R}$.

Proof. The proof is standard, however, details are quite tedious. For the reader convenience, we present a detailed proof proposed by M. Drmota in Appendix A. Here we only sketch the main idea. We first proof (29) for characteristic functions $\chi_I(x_k)$. This follows from Definition 1. Then, we approximate f by a step function (i.e., a combination of characteristic functions) and use the definition of the Riemann integral to bound the integral from below and above. We show that when $n \to \infty$ these bounds coincide with the left-hand side of (29). ■

Remark. A stronger version of Theorem 4 is possible. It can be proved (cf. Drmota [6]) that for functions $f : [0,1] \to \mathbb{R}$ with bounded variations $V(f)$ we have

$$\begin{aligned}
&\left| \lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle \alpha k + \beta n \rangle) - \int_0^1 f(t)\, dt \right| \\
\leq\ & V(f) \cdot \inf_{\eta>0} \left(e^{-C\eta^2} + \eta \max_{\eta\sqrt{n} \leq N \leq 2\eta\sqrt{n}} D_N(\alpha k) \right)
\end{aligned}$$

where $D_N(\alpha k)$ is the discrepancy of the sequence $\langle \alpha k \rangle_{k=1}^N$. The discrepancy of the Weyl sequence $\langle \alpha k \rangle$ has been extensively studied and is well understood (cf. [7, 16]). □

To use Theorem 4 effectively, one needs a simple criterion to verify whether a sequence is B-u.d. mod 1. Such a criterion, fortunately, exists and it is basically due to Weyl. Before we formulate it, we first observe that we can relax the condition of Theorem 4 to functions f that are continuous with period 1.

Theorem 5 (Weyl's Criterion) *A sequence x_n is B-u.d. mod 1 if and only if*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{2\pi i m x_k} = 0 \tag{30}$$

holds for all $m \in \mathbb{Z} - \{0\}$.

Proof. The proof again is standard and the reader is referred to textbooks such as [7, 16] (cf. Chapter 8 of [28] for a brief discussion). Basically, it is based on the fact that by Weierstrass's *approximation theorem* every Riemann integrable function f of period 1 can be uniformly approximated by a trigonometric polynomial (i.e., a finite combination of functions of the type $e^{2\pi i m x}$). ■

Now, we are in position to finish our derivation. We first prove that $\langle \alpha k \rangle$ is B-u.d. mod 1. Indeed, by Weyl's criterion

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{2\pi i m (k\alpha)} = \lim_{n\to 0} \left(p e^{2\pi i m \alpha} + q \right)^n = 0 \tag{31}$$

provided α is irrational. Hence, by Theorem 4, with $f(t) = t$ and $y = \beta n$, we immediately obtain

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle = \int_0^1 t dt = \frac{1}{2}. \tag{32}$$

This proves the irrational part of Theorem 2.

3.2 Rational Case

Now, we turn our attention to the case when α is rational. We assume $\alpha = M/N$ where M, N are integers such that $\gcd(N, M) = 1$. Let $p_{n,k} = \binom{n}{k} p^k q^{n-k}$ and we denote by S_n the sum in (16). We proceed as follows

$$\begin{aligned} S_n &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left\langle k\frac{N}{M} + \beta n \right\rangle = \sum_{k=0}^{n} \binom{n}{k} p_{n,k} \left\langle k\frac{N}{M} + \beta n \right\rangle \\ &= \sum_{\ell=0}^{M-1} \sum_{m:\ k=\ell+mM\le n} p_{n,k} \left\langle \ell\frac{N}{M} + N + \beta n \right\rangle \\ &= \sum_{\ell=0}^{M-1} \sum_{m:\ k=\ell+mM\le n} p_{n,k} \left\langle \frac{\ell}{M} + \beta n \right\rangle \\ &= \sum_{\ell=0}^{M-1} \left\langle \frac{\ell}{M} + \beta n \right\rangle \sum_{m:\ k=\ell+mM\le n} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned} \tag{35}$$

To evaluate the last sum we need the following simple lemma. It asserts that if one picks every Mth term of the binomial distribution, then the total probability of this sample is "well" approximated by $1/M$.

Lemma 1 *For fixed $\ell \le M$ and M, there exist $\rho < 1$ such that*

$$\sum_{m:\ k=\ell+mM\le n} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{M} + O(\rho^n). \tag{36}$$

Proof. Let $\omega_k = e^{2\pi i k/M}$ for $k = 0, 1, \ldots, M-1$ be the Mth root of unity. It is well known that (cf. [28])

$$\frac{1}{M} \sum_{k=0}^{M-1} \omega_k^n = \begin{cases} 1 & \text{if } M|n \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

where $M|n$ means that M divides n. In view of this, we can write

$$\begin{aligned} \sum_{m:\ k=\ell+mM\le n} \binom{n}{k} p^k q^{n-k} &= \frac{1 + (p\omega_1 + q)^{n-\ell} + \cdots + (p\omega_{M-1} + q)^{n-\ell}}{M} \\ &= \frac{1}{M} + O(\rho^n), \end{aligned} \tag{38}$$

since $|(p\omega_r + q)| = p^2 + q^2 + 2pq\cos(2\pi r/M) < 1$ for $r \neq 0$. This proves the lemma. ■

From now on, we only deal with the sum $S_n = \frac{1}{M}\sum_{\ell=0}^{M-1}\left\langle \frac{\ell}{M} + \beta n\right\rangle$ of (35) ignoring the error term $O(\rho^n)$. We use again the Fourier series (26) and (37) to obtain

$$\begin{aligned}
S_n &= \frac{1}{M}\sum_{\ell=0}^{M-1}\left(\frac{1}{2} - \sum_{m\neq 0} c_m e^{2\pi i m(\ell/M+\beta n)}\right) \\
&= \frac{1}{2} - \sum_{m\neq 0} c_m e^{2\pi i m n\beta}\frac{1}{M}\sum_{\ell=0}^{M-1} e^{2\pi i m\frac{\ell}{M}} \\
&\overset{(37)}{=} \frac{1}{2} - \frac{1}{M}\sum_{m=kM\neq 0} c_{kM} e^{2\pi i k M\beta n} \\
&\overset{(26)}{=} \frac{1}{2} - \frac{1}{M}\left(\frac{1}{2} - \langle \beta n M\rangle\right). \qquad (39)
\end{aligned}$$

This proves the rational part of (17), and completes the proof of Theorem 2.

4 Derivation of the Huffman Code Redundancy

We establish here Theorem 1, that is, we shall prove (10). Due to (15), we must only deal with the asymptotics of the following sum

$$T_n = \sum_{k=0}^{n}\binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n\rangle}. \qquad (40)$$

As in the previous section, we shall apply Fourier series, but this time we need the following for some $a > 0$

$$2^{-\langle x\rangle/a} = C_0(a) + \sum_{m\neq 0} C_m(a) e^{2\pi i m x}, \qquad (41)$$

where

$$\begin{aligned}
C_0(a) &= \frac{a}{\ln 2}\left(1 - 2^{-1/a}\right), \qquad (42)\\
C_m(a) &= \frac{a}{2\pi i m a + \ln 2}\left(1 - 2^{-1/a}\right), \qquad m \neq 0. \qquad (43)
\end{aligned}$$

4.1 Irrational Case

We first consider α irrational. To analyze T_n we can directly applied Theorem 4 with $f(t) = 2^{-t}$ and $y = \beta n$. This leads to

$$\lim_{n\to\infty} T_n = \int_0^1 2^{-t} dt = \frac{1}{2\ln 2} \qquad (44)$$

which proves Theorem 1 for α irrational.

4.2 Rational Case

Next, we consider α rational. This turns out to be a more intricate case. First of all, applying Lemma 1 we can re-write T_n when $\alpha = N/M$ $(\gcd(N, M) = 1)$ as

$$T_n = \frac{1}{M} \sum_{\ell=0}^{M-1} 2^{-\langle \frac{\ell}{M} + \beta n \rangle} + O(\rho^n) \tag{45}$$

for some $\rho < 1$. Ignoring the error term, we proceed as follows

$$\begin{aligned}
T_n &= C_0(1) + \frac{1}{M} \sum_{m \neq 0} C_m(1) \sum_{\ell=0}^{M-1} \exp\left(2\pi i m \left(\frac{\ell}{M} + n\beta\right)\right) \\
&= C_0(1) + \sum_{m \neq 0} C_m(1) e^{2\pi i m n \beta} \frac{1}{M} \sum_{\ell=0}^{M-1} \left(e^{2\pi i \ell / M}\right)^m \\
&\overset{(37)}{=} C_0(1) + \sum_{m = kM \neq 0} C_{kM}(1) e^{2\pi i n \beta k M} \qquad (46) \\
&= \frac{1}{2\ln 2} + \frac{1}{M} \sum_{k \neq 0} \frac{1}{2(2\pi i k + \ln 2/M)} e^{2\pi i k n \beta M} \\
&= \frac{1}{2\ln 2} + \frac{1}{2(1 - 2^{-1/M})} \sum_{k \neq 0} \frac{(1 - 2^{-1/M})}{(2\pi i k + \ln 2/M)} e^{2\pi i k (n \beta M)} \\
&\overset{(41)}{=} \frac{1}{2\ln 2} + \frac{1}{2(1 - 2^{-1/M})} \sum_{k \neq 0} C_k(M) e^{2\pi i k (n \beta M)} \\
&= \frac{1}{2M(1 - 2^{-1/M})} 2^{-\langle n \beta M \rangle / M}. \qquad (47)
\end{aligned}$$

This establishes the rational part of (11), and completes the proof of Theorem 1.

In passing, we should point out that our proofs of the rational cases for the Shannon code and the Huffman code can be somewhat simplified by the following theorem (suggested by M. Drmota) that completely avoids Fourier analysis. The proof is is elementary and can be found in Chapter 8 of [28].

Theorem 6 *Let $0 < p < 1$ be a fixed real number and suppose that $\alpha = \frac{N}{M}$ is a rational number with $\gcd(N, M) = 1$. Then, for every bounded function $f : [0, 1] \to \mathbb{R}$ we have*

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle k\alpha + y \rangle) = \frac{1}{M} \sum_{l=0}^{M-1} f\left(\frac{l}{M} + \frac{\langle My \rangle}{M}\right) + O(\rho^n) \tag{48}$$

uniformly for all $y \in \mathbb{R}$ and some $\rho < 1$.

5 Derivation of the Golomb Code Redundancy

We assume that $\theta^\ell = 1/2$ as in (20), and estimate the entropy $H(\theta)$ (cf. (22)) as $\ell \to \infty$ (i.e, $\theta = 2^{-1/\ell} \to 1$). Observe that

$$H(\theta) = -\lim_{\ell\to\infty} \log_2(1 - 2^{-1/\ell}) + \log_2 e. \tag{49}$$

Using the following Taylor's expansion

$$\log_2(1 - 2^{-x}) = \log_2(\ln(2) + \log_2(x) - \frac{1}{2}x + \frac{\ln(2)}{24}x^2 + O(x^3), \qquad x \to 0, \tag{50}$$

we arrive at

$$H(\theta) = \log_2 \ell + \log_2(\log_2 e) + \frac{1}{2\ell} + O(\ell^{-2}). \tag{51}$$

The average redundancy $\bar{R}^G_\ell = \mathbf{E}[L] - H(\theta)$, where $\mathbf{E}[L]$ is given by (20), follows after some simple algebra

$$\bar{R}^G_\ell = 1 - \langle \log_2 \ell \rangle + 4 \cdot 2^{-2^{1-\langle \log_2 \ell \rangle}} - \log_2(\log_2 e) - \log_2 e - \frac{1}{2\ell} + O(\ell^{-2}) \tag{52}$$

as $\ell \to \infty$ or $\theta \to 1$. As before $\langle x \rangle$ denotes the fractional part of x. This proves (23) of Theorem 3. We should also point out that using the above approach we can get a full asymptotic expansion of $\bar{R}^G_\ell$ as $\ell \to \infty$.

By Weyl's criterion (cf. Theorem 5) $\langle \log_2 n \rangle$ is *not* uniformly distributed mod 1 (but it is dense in $(0,1)$). Similar conclusion holds for $2^{-2^{1-\langle \log_2 \ell \rangle}}$. Thus, by (52) the average redundancy $\bar{R}^G_\ell$ asymptotically oscillates within a certain interval without reaching a limit, as observed in Figure 3.

To compute the magnitude of the oscillation, let us define $C = \log_2(\log_2 e) + \log_2 e = 1.971461414\ldots$. We set

$$g(x) = 1 - x + 4 \cdot 2^{-2^{1-x}} - C \tag{53}$$

for $0 \le x \le 1$. (Observe that $g(x)$ is equal to R^G_ℓ when $x = \langle \log_2 n \rangle$.) One derives that $g(x)$ achieves it maximum value $g(x_1) := \max\{g(x)\} = .327344112\ldots$ at

$$x_1 = \log_2 \left(\frac{-2\ln(2)}{W(-0.25 \log_2 e)} \right) \tag{54}$$

where $W(x)$ is the Lambert-W function defined as $W(x)e^{W(x)} = x$ (cf. [4]). Similarly, the function $g(x)$ achieves its minimum value $g(x_2) := \min\{g(x)\} = .251005712\ldots$ at

$$x_2 = \log_2 \left(\frac{-2\ln(2)}{W(-1, -0.25 \log_2 e)} \right) \tag{55}$$

where $W(-1,x)$ is a branch of the Lambert-W function (cf. [4]). It is also easy to see that the average redundancy oscillates around $g(0) = g(1) = 2 - C = .0285385862\ldots$. This proves Theorem 3.

Appendix A: Drmota's Proof of Theorem 4

Let $\mathcal{L}$ be the linear space of all (Lebesgue) integrable functions $f : [0,1]^k \to \mathbb{C}$ satisfying

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}} \left| \sum_{k\geq 0} p_{n,k} f(\langle x_k + y\rangle) - \int_0^1 f(t)\,dt \right| = 0. \tag{A.1}$$

Of course, $\mathcal{L}$ contains all constant functions, i.e. $\mathcal{L} \neq \emptyset$. Next we prove the following closure property of $\mathcal{L}$.

> **Fact A**. Suppose $f : [0,1] \to \mathbb{R}$ is an integrable function such that for every $\varepsilon > 0$ there exist functions $g_1, g_2 \in \mathcal{L}$ with $g_1 \leq f \leq g_2$ and
>
> $$\int_0^1 (g_2(t) - g_1(t))\,dt < \varepsilon.$$
>
> Then $f \in \mathcal{L}$.

For this purpose, let $m_n(f,y)$ denote the positive linear functionals

$$m_n(f,y) := \sum_{k\geq 0} p_{n,k} f(\langle x_k + y\rangle)$$

and $m(f)$ the positive linear functional

$$m(f) := \int_0^1 f(t)\,dt.$$

By $m_n(g_1,y) \leq m_n(f,y) \leq m_n(g_2,y)$ and $m(g_1) \leq m(f) \leq m(g_2)$ we immediately get

$$\begin{aligned} m(g_1) &= \liminf_{n\to\infty} m_n(g_1,y) \leq \liminf_{n\to\infty} m_n(f,y) \\ &\leq \limsup_{n\to\infty} m_n(f,y) \leq \limsup_{n\to\infty} m_n(g_2,y) \\ &= m(g_2) \end{aligned}$$

that implies

$$|m(f) - \liminf_{n\to\infty} m_n(f,y)| < \varepsilon$$

and

$$|m(f) - \limsup_{n\to\infty} m_n(f,y)| < \varepsilon$$

for every $\varepsilon > 0$. Thus

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}} |m_n(f,y) - m(f)| = 0,$$

and $f \in \mathcal{L}$.

Next, suppose that $f : [0,1] \to \mathbb{R}$ is a continuous function. Then by the Weierstrass approximation theorem, for every $\varepsilon > 0$ there exist two trigonometric polynomials g_1, g_2 with $g_1 \le f \le g_2$ and

$$\int_0^1 (g_2(t) - g_1(t))\,dt < \varepsilon.$$

By fact A, all continuous functions $f : [0,1] \to \mathbb{R}$ are contained in $\mathcal{L}$.

Now, it is easy to derive that every characteristic function $f = I_A$ of any interval $A \subseteq [0,1]$ is in $\mathcal{L}$, too. It is obvious, that for every $\varepsilon > 0$ there exist two continuous functions g_1, g_2 with $g_1 \le I_A \le g_2$ and

$$\int_0^1 (g_2(t) - g_1(t))\,dt < \varepsilon.$$

Again by Fact A all step functions (i.e. finite linear combinations of characteristic functions) are in $\mathcal{L}$.

Finally, if $f : [0,1] \to \mathbb{R}$ is a Riemann integrable function, then by definition for every $\varepsilon > 0$ there exist two step functions g_1, g_2 with $g_1 \le f \le g_2$ and

$$\int_0^1 (g_2(t) - g_1(t))\,dt < \varepsilon,$$

and by Fact A this completes the proof.

ACKNOWLEDGMENT

This work would be impossible without many happy occurrences. John Kieffer asked me about three years ago if "analytic information theory" can handle the Huffman code redundancy. When working with Philippe Jacquet on the entropy computations via analytic methods (cf. [15], and with H-K. Hwang on the redundancy of CTW code (yet unpublished), I understood that the theory of *sequences distributed modulo* 1 is relevant, and I did some initial work on the Shannon code. Then, happily enough, Julia Abrahams, after my talk at Princeton during the *Fourth Seminar on Analysis of Algorithms*, sent Stubley's paper [25] that prompted me to look again at the Huffman redundancy problem. Michael Drmota helped me with many technical questions concerning uniform distribution of sequences, supplied a detailed proof of Theorem 4, and provided many insightful comments. Svante Janson had additional comments during my talk at the *Fifth Seminar on Analysis of Algorithms*, Barcelona, June 1999. P. Stubley generously shared his unpublished paper [26]. Finally, during *LATIN*'2000 in beautiful Punta del Este G. Seroussi asked me to look at the redundancy of Golomb's code. N. Merhav had many useful comments and suggested to include the redundancy of the Golomb code in this paper. To all of them I thank whole-heartily.

References

[1] J. Abrahams, "Code and Parse Trees for Lossless Source Encoding, *Proc. of Compression and Complexity of SEQUENCE'97*, Positano, IEEE Press, 145–171, 1998.

[2] R. Capocelli and A. de Santis, Tight Upper bounds on the Redundancy of Huffman Codes, *IEEE Trans. Information Theory*, 35, 1084-1091, 1989.

[3] R. Capocelli and A. de Santis, New Bounds on the Redundancy of Huffman Codes, *IEEE Trans. Information Theory*, 37, 1095-1104, 1991.

[4] R. Corless, G. Gonnet, D. Hare, D. Jeffrey and D. Knuth, On the Lambert W Function, *Adv. Computational Mathematics*, 5, 329–359, 1996.

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

[6] M. Drmota, Private Communications.

[7] M. Drmota and R. Tichy, *Sequences, Discrepancies, and Applications*, Springer Verlag, Berlin Heidelberg 1997

[8] R. Gallager, *Information Theory and Reliable Communications*, New York, Wiley 1968.

[9] R. Gallager, Variations on the Theme by Huffman, *IEEE Trans. Information Theory*, 24, 668-674, 1978.

[10] R. Gallager and D. van Voorhis, Optimal Source Codes for Geometrically Distributed Integer Alphabets, *IEEE Trans. Information Theory*, 21, 228-230, 1975.

[11] S. Golomb, Run-Length Coding, *IEEE Trans. Information Theory*, 12, 399-401, 1966.

[12] R. M. Gray, Quantization Noise Spectra, *IEEE Trans. Information Theory*, 36, 1220-1244, 1990.

[13] D. Huffman, A Method for the Construction of Minimum Redundancy Codes, *Proc. IRE*, 40, 1089-1101, 1952.

[14] P. Jacquet and W. Szpankowski, Asymptotic Behavior of the Lempel-Ziv Parsing Scheme and Digital Search Trees, *Theoretical Computer Science*, 144, 161-197, 1995.

[15] P. Jacquet and W. Szpankowski, Entropy Computations via Analytic Depoissonization, *IEEE Trans. Information Theory*, 45, 1072-1081, 1999.

[16] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. John Wiley, New York 1974.

[17] G. Louchard and W. Szpankowski, On the Average Redundancy Rate of the Lempel-Ziv Code, *IEEE Trans. Information Theory*, 43, 2–8, 1997.

[18] D. Manstetten, Tight Upper Bounds on the Redundancy of Huffman Codes, *IEEE Trans. Information Theory*, 38, 220-22, 1980.

[19] N. Merhav, G. Seroussi and M. Weinberger, Optimal Prefix Codes for Sources with Two-Sided Geometric Distributions, *IEEE Trans. Information Theory*, 46, 121-135, 2000.

[20] F. Olver, *Asymptotics and Special Functions*, Academic Press, New York 1974.

[21] J. Rissanen, Complexity of Strings in the Class of Markov Sources, *IEEE Trans. Information Theory*, 30, 526–532, 1984.

[22] S. Savari, Variable-to-Fixed Length Codes for Predictable Sources, *Proc. Data Compression Conference*, 481-490, Snowbird 1998.

[23] S. Savari and R. Gallager, Generalized Tunstall Codes for Sources with Memory, *IEEE Trans. Information Theory*, 43, 658-668, 1997.

[24] Y. Shtarkov, Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, 23, 175–186, 1987.

[25] P. Stubley, On the Redundancy of Optimum Fixed-to-Variable Length Codes, *Proc. Data Compression Conference*, 90-97, Snowbird 1994.

[26] P. Stubley, The Redundancy of Optimum Codes for Multinomially Distributed Sources, preprint 1998.

[27] W. Szpankowski, On Asymptotics of Certain Recurrences Arising in Universal Coding, *Problems of Information Transmission*, 34, No.2, 55-61, 1998.

[28] W. Szpankowski, *Average Case Analysis of Algorithms in Sequences*, John Wiley & Sons, New York, 2000.

[29] M. Weinberger, N. Merhav, and M. Feder, Optimal Sequential Probability Assignments for Individual Sequences, *IEEE Trans. Information Theory*, 40, 384–396, 1994.

[30] F.M. Willems, Y. Shtarkov and T. Tjalkens, The Context–Tree Weighting Method: Basic Properties, *IEEE Trans. Information Theory*, 41, 653–664, 1995.

[31] J. Ziv, Back from Infinity: A Constrained Resources Approach to Information Theory, *IEEE Information Theory Society Newsletter*, 48, 30-33, 1998.

[32] A. Zygmund, *Trigonometric Series*, Cambridge University Press, New York, 1959.